

研究方法·基础应用

调查研究实操指导：
细节与示例

HOW TO CONDUCT SURVEYS: A STEP-BY-STEP GUIDE 5TH EDITION

阿琳·芬克（Arlene Fink）著

齐 心 译





DIAOCHA YANJIU SHICAO ZHIDAO XIJIE YU SHILI

调查研究实操中的注意事项：

- 问卷的长度由你需要了解什么，需要提出多少问题才能使相应的结果可信，以及被调查者能拿出多少时间和注意力用于调查来决定。
- 问卷中的问题应该从最熟悉的到最不熟悉的进行排列。
- 把相对容易回答的问题放在问卷的末尾。
- 好的电话访谈和面对面访谈的关键是培训。
- 应该对访谈予以监控，以确保访谈员以相同的方式提出问题，并使回答的一致性最大化。



定价：32.00元

调查研究实操指导：细节与示例

HOW TO CONDUCT SURVEYS:
A STEP-BY-STEP GUIDE
5TH EDITION

阿琳·芬克（Arlene Fink）著
齐　心 译

重庆大学出版社

图书在版编目(CIP)数据

调查研究实操指导:细节与示例/(美)芬克(Fink, A.G.)著;齐心译.—重庆:重庆大学出版社, 2016.3

(万卷方法)

书名原文:How to Conduct Surveys: A Step-by-Step Guide

ISBN 978-7-5624-9703-5

Ⅰ.①调… Ⅱ.①芬…②齐… Ⅲ.①调查研究—方法 Ⅳ.①C31

中国版本图书馆 CIP 数据核字(2016)第 045282 号

调查研究实操指导:细节与示例

阿琳·芬克 著

齐 心 译

策划编辑:雷少波 林佳木 邹 荣

责任编辑:林佳木 邹 荣 版式设计:林佳木

责任校对:秦巴达 责任印制:赵 晟

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023)88617190 88617185(中小学)

传真:(023)88617186 88617166

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn(营销中心)

全国新华书店经销

重庆华林天美印务有限公司印刷

*

开本:940mm×1360mm 1/32 印张:7.25 字数:202 千

2016 年 4 月第 1 版 2016 年 4 月第 1 次印刷

ISBN 978-7-5624-9703-5 定价:32.00 元

关于作者

阿琳·芬克(博士)是加利福尼亚大学洛杉矶分校医学和公共卫生教授以及兰利研究所所长。她的主要兴趣包括评估和调查研究,以及开展研究文献综述和对文献质量进行评估。芬克博士已经在医学、公共卫生和教育领域开展过很多项研究。她是加州大学洛杉矶分校的罗伯特·伍德·约翰逊临床学者计划的教师,同时还是该校赌博研究及 IMPACT(促进加州癌症晚期病人的接近、咨询和治疗)项目的顾问。她为国内国际的许多机构提供咨询,如法国巴黎的 IPPSA 以及澳大利亚的维多利亚半岛健康组织。芬克教授频繁在全球多地讲课或演讲,她是 130 多篇同行评议文章和 15 种教科书的作者。

前言

本书第5版与前4版的目标相同：指导读者自行开展严谨的调查，或评估其他人所作调查的可信性和有用性。与之前的版本一样，本版也给出了达到目标的各步骤的实用指导。重在如何选择合适的调查类型、撰写调查问题和选项、格式化调查、确定应该纳入的被调查者的特征和人数、应该对他们调查多少次，以及分析和报告调查结果。

本版与前版遵循的基本观点也是相同的：强调在按下“发送”、寄出第一份问卷或开始第一个访谈之前，进行细致的规划和测试的必要性。但是各版的相似性也仅止于此，第5版中的每一章都有一些你在其他版中找不到的东西。

第4版曾经预言基于网络采用笔记本电脑和台式机、电话和平板电脑的调查将会变得更为重要。事实也是如此，本版反映了这种变化。以下是第5版中一些新主题的概述。你将看到对互联网的特别强调。

- 何时使用网上调查（你需要覆盖广阔的地理范围；你有正确的电子邮件地址；人们有高网速）
- 网上调查的特有问题（你需要有效的电子邮件地址去通知人们你的调查。许多人有多个电子邮件地址，所以你应该选择哪个？）
- 如何选择网上调查软件（你是否应该只使用大品牌？免费的情况下你能对调查模板做多少改变？你能在其“消失”在网络空

间前保持调查多长时间?)

- 网上调查独有的伦理考虑(被调查者与调查网站之间通讯的伦理议题;网站与调查者之间通讯的伦理议题;调查组织的隐私政策的最低标准。)
- 如何格式化网络调查,包括提供使用浏览器或调查页面的前进和后退按钮的指导语;进度条;确定调查使用连续的文档还是划分成每屏一或多个问题;如何选择问题和回答形式
- 调查修辞:如何检查调查的可读性;良好修辞和调查用语
- 如何在网上和文献中查找和评估调查,以便你改编它们或从中找出能够满足你的需求的具体问题
- 对可读性和有效性的新思考
- 样本量、效应量和统计效力方面的更新
- 加权:设计加权和无回答加权
- 如何设计海报描述你的调查目标、方法和发现

第 5 版在形式上也有所更新。我们放大了字号以利阅读,并且使用更高端的软件重做了图形。

虽然第 5 版有很多新东西,但也保留了前版很多内容,因为读者和评论者认为这些内容有用。这包括大量例子和清单,以及每章末尾的帮助读者掌握最重要概念的"小结"。总结尽管简单,但我认为这是很有价值的,因为它帮助人们了解最应该关注的基本内容。为了加强对各要点的理解,每章都有一些练习题,其中很多都是第 5 版新增的。这些练习会提供给你一些调查报告和研究,使你能够检验自己理解和评估它们的技能。我还扩展和更新了参考文献,并附加了一些网站、调查报告和包含调查的研究。我列出了网址,当你需要从其他调查改编问题的时候可以访问这些网站。总得来说,调查者在整个调查过程中必须不断选择。有时候选择是困难的,为此全书都可找到"做出决定"这样的小节。这些小节的目的是帮助读者在每个关键节点权衡利弊做出明智的选择。你能够得到选择调查类型、问题类型、抽样策略、研究设计和数据分析方法等方面的指导。

我尽量使本书适合于每个需要了解如何做调查的人,无论其背景如何。这本书实践的内容多理论的内容少。它的目标是帮助你做调查。如果你希望了解调查理论,请借助本书的文献目录。我认为本书适合自学,也可在讲习班、研讨班和正式的课堂上使用。

从个人角度来说,我感谢我的编辑维姬·奈特(Vicki Knight)在整个编写过程中给予我的支持。她见证了我们邀请的众多我深怀感激的评论者:

金妍秀(Yeonsoo Kim) 内华达州拉斯维加斯大学
卡拉·米勒(Carla Mueller) 圣弗朗西斯大学
理查德·J.哈瑞斯(Richard J.Harris) 德克萨斯大学圣安东尼奥分校
黛博拉·唐森(Deborah Townsend) 孟菲斯大学
安德斯·汉森(Anders Hansen) 莱斯特大学
上桥川崎·奥康纳(Nahoko Kawakyu O'connor) 罗切斯特大学
乔赛特·贝塔尼-萨提可夫(Josette Bettany-Saltikov) 提兹塞德大学
娜西·M.桑德斯(Nacy M.sanders) 华盛顿州立大学温哥华分校
露丝·诺曼(Ruth T.Norman) 威尔明顿大学
布莱恩·杜斯比伯(Brian Dusbiber) 圣凯瑟琳大学
罗斯娜·麦卡利(Roseanna McCleary) 贝克斯菲尔德加州州立大学
辛西亚·J. 麦格雷格(Cynthia J. MacGregor) 密苏里州立大学
卡琳娜·M. 施莱弗勒(Karina M. Shreffler) 俄克拉何马州立大学
马克·S. 罗森鲍姆(Mark S. Rosenbaum) 北伊利诺斯大学

他们的建议是极有价值的,我真心感谢他们为此所付出的时间和努力,所以我确实尽力吸纳他们的建议。我确信这些建议对本书大有益处,有时候我没采纳仅仅是为了偷点儿小懒。我还要感谢 Sage 的工作团队为本书重定版式和改写提供的帮助,是他们的努力让本书变得更好。



目录 Contents

1 做调查：每个人都在做的事

2 调查问卷：问题、量表和格式

3 调查研究的一些实践要点

4 调查抽样

1

做调查:每个人都在做的事



调查无处不在。你可以在医院、学校、飞机和旅馆房间找到它们。调查被用来收集关于人们如何描述、比较或解释他们的知识、感觉、价值和行为的信息。通常调查采取自填问卷和访谈的形式。自填问卷可以手填("纸和笔")或者网填(在电脑和平板上基于网络的方式)。访谈则可以当面("面对面")或通过电话(固定电话和移动电话)进行。

调查数据被各个领域(包括商业、健康、教育、社会福利和政治)中的项目规划者、评估者、监督者、管理者、研究者和政策领导者所用。这些数据之所以有用是因为它们获取的信息直接来自人。

调查者必须决定调查的总体目的和具体问题。他们还必须知道需要接触什么人以及多少人(抽样),何时和以多高频率开展调查(设计)。调查者还必须处理、分析和解释数据。

调查者要在调查类型(自填问卷或访谈)和执行方法(邮件、电话或网络)之间做出选择,需要寻找最有可能产生可信且准确结果的组合,并且在希望的调查类型和执行方法与现有资源之间做出权衡。

调查目的和方法处于一个连续统中。有些调查可能具有深远和普遍的影响,因而其方法必须是科学的。由美国政府主持的人口健康调查是科学调查的一个例子。其他调查的开展可能是为了满足某些具体需求,其方法不一定总是能达到最严格的科学标准,但是也必须能够带来准确的结果,因而必须使用可靠且有效的技术。调查某所学校学生的暑期阅读选择以确保图书馆做好准备,这就是一个调查用来满足某一具体需求的例子。

调查者必须注重保护被调查者的隐私,并对回答保密。大多数的机构都有保护"人类被试"的规则以确保他们同意作答是属于知情同意的。网上调查在保密方面有特殊规则,因为这类调查易受外部攻击,因而需要保护三类通信关系:调查者与回答者、回答者与网络服务器以及网络服务器与调查者。

什么是调查?

调查是一种信息收集方法,用来描述、比较或解释个人性和社会性的知识、感受、价值、偏好和行为。调查可以由个人独自或在导语指引下自填完成,也可以是当面或通过电话被访谈完成。有些调查是纸质或网页的,被调查者或在家里独立完成或在一个集中地点(比如一个健康中心)完成,被调查者或传统邮件或网上提交所完成的调查。调查可以是互动式的,通过问题引导被调查者。互动式调查还可以提供音视频帮助信息。

为什么开展调查至少有三个好理由。

理由1:需要制定一项政策或规划一个项目

■ 示例　　**满足政策或项目需求的调查**

- YMC 公司希望确定每天的营业时间。这家公司对员工进行调查以了解他们希望在什么样的8小时时段里工作。
- 健康志愿协会全国办公室考虑为其雇员提供儿童日间照料。有多少人有幼童? 多少人会使用相关设施?

- 十年前,巴特利学区改变了其语言艺术课程。从那时起就有一些人认为这一课程已经过时了。英语教师们是怎样认为的?如果需要修改,应该改成什么样子呢? ■

理由 2:你希望评估一个旨在改变人们的知识、态度、健康或福利的项目的效果

■ 示例　　**项目评估中的调查**

- YMC 公司设立了两个帮助人们了解在非常规时间工作的优缺点的项目。一个项目采取个人咨询和自行观看特制 DVD 的形式。第二个项目在较大的群体中开展。在每个项目完成后的第 6 个月开展了一项调查,以了解员工是否认为他们已经获得了所需的信息。调查还意图了解他们是否会推荐其他人参与同一个项目,以及他们对其工作时间表的满意度。
- 健康志愿协会尝试了两种托儿方式。一种主要是"以孩子为中心的",通常由孩子们决定他们在项目期间最喜欢做的活动;另一种偏于学业和艺术,每天定时教授孩子们阅读、乐器演奏和跳舞。从积极参与并喜欢课程内容的家长、孩子和员工的角度看,哪个项目最令人满意?协会调查了家长、孩子和员工以寻求答案。
- 巴特利学区改变了其语言艺术课程。开展了一项调查以了解这种改变是否以及如何影响了父母和学生对高中课程的意见。 ■

理由 3:你是一位研究者,通过调查获取如何指导研究和项目的信息

■ 示例　　**为了研究的调查**

- 因为 YMC 公司有很多培训项目,所以他们希望研究如何更好开展成人教育。人们是喜欢自学还是正规的课堂?文本材料合适还是电影和 DVD 更佳?他们对计算机辅助学习或直接通过网络学习的感受如何?作为研究的一部分,也是为了确保包括所有的可能性,这家公司对员工进行抽样调查以了解他们的偏好。
- 健康志愿协会打算与一所地方大学合作开展一项关于学前教育的研究。该协会对参与新的日托项目的家长进行了调查。调查询问了参加者的受教育水平和收入。研究者需要此类数据检验他们的一个主要假设——家

长的受教育水平和收入越高就越有可能选择两个学前项目中更偏向学业的那个。

- 巴特利学区参与了一项美国政府资助的英语教学全国性研究。该研究的研究者假设课堂教学更多的取决于教师的教育背景和阅读偏好,而非正规的课程大纲。开展一项调查以了解教师的教育背景和阅读习惯,以便利用这些数据检验研究者的假设。 ■

最好何时调查?

有许多方法可以获取关于人的信息,调查只是其中之一。考察一家主要目标是为社区提供多样化的服务的青年中心。它向年龄在12到21岁之间的城市居民(无论其背景如何)提供医疗、财务、法律和教育援助。这个项目最引以为傲的是它的集成方法,认为中心的效率来自于在同一地点向所有参加人提供服务。由于中心已有10年的历史,所以打算开展一项调查以了解其究竟有多成功。参与者和员工是否满意?年轻人使用了哪些服务?中心是否确实是复合服务型的?参与中心是否增进了人们的健康和其他需求的满足?调查决定采用邮寄自填问卷方式以回答这些问题和其他一些问题。以下是问卷的一些节选。

■ 示例　**一份过于雄心勃勃的自填问卷中的问题**

5.你的血压现在正常吗?

正常　1

不正常　2

7.最近12月内你是否使用过下列社会服务?(请针对每项服务分别作答)

服　务	1.是	2.否
医疗……		
法律……		
财务……		
教育……		

10.你对下列服务的满意程度如何?请针对每项服务分别作答

服　务	1.非常满意	2.满意	3.中等	4.不满意	5.很不满意
日常咨询会议	5	4	3	2	1
法律援助设施	5	4	3	2	1
图书馆	5	4	3	2	1

11.在一个5分钟的时段内有多长时间是医生在听你说(而不是医生对你说)(请选择一个答案)?

□ 少于1分钟

□ 大约1到2分钟

□ 超过2分钟

问卷被送到一位评阅人那里,他建议去掉问题5、7和11,只保留问题10。评阅人认为对某些类型的信息而言调查并非最佳获取途径。理由如下:

问题5要求报告个人的血压。血压是否正常?通常,这类信息从其他渠道(比如医疗记录)获取会更为准确。许多人可能难于回忆起自己的血压,也不大会界定“正常”的血压。

如果你确信人们的回忆是准确的,则问题7是可以的。否则,中心的记录可能是获取这类服务使用信息的更好渠道。

问题11要求病人告知医生用于倾听而不是交谈的时间。如果你对病人的感知感兴趣,那么这个问题是不错的。不过,如果你希望获取医生用于倾听而不是和病人交谈的准确时间,由一个客观的旁观者进行观察可能更为可取。

问题10是恰当的。只有参与者才能告诉你他们的满意程度如何。除此之外别无他途。

对于决策而言,调查绝非唯一的信息渠道,也不一定总是最适用的。以下是其他一些信息渠道:

- 观察或目击报告;摄影、录像和录音资料
- 表现测评,要求某个人执行某项任务(如给一个班级授课);或

观察者评价他人表现情况

- 能力或知识的书面考试

- 依托现有文献的记录查阅,例如对医疗和入学记录的查阅;对已发表和未发表的文章和日志或者记录和拍摄下来的文件的内容的分析。

决定政策或规划,评估项目和开展研究时,如果你需要的信息必须直接来自于人,就可以使用调查方法。他们提供的数据是关于感觉和知觉、价值、习惯和个人背景,或者诸如年龄、健康、教育和收入等人口学特征的描述。

有时候调查是与其他信息渠道结合在一起的。进行评估和研究时尤其如此。

■ 示例　　**与其他信息渠道相结合的调查**

- 作为其托儿项目评估的一部分,健康志愿协会调查了父母、孩子和员工的参与和满意程度。此外,协会还查阅了财务记录以评估每个项目的成本,并进行了标准化测验以评估孩子们的入学准备情况。
- YMC 公司正在研究成人是如何学习的。定期开展了成就和表现测验。此外,一项调查提供关于成人喜爱学习程度的补充数据。 ■

自填问卷和访谈是调查的关键

所有调查都包括:(1)问题和回答。为了获取准确数据,你必须说明调查的(2)抽样和设计,(3)数据处理或"管理"和分析,(4)小规模测试,以及(5)回答率。调查结果以书面(印在纸上或电子复制)和口头报告的形式呈递。

问题和回答

来自调查的信息是通过提问获取的。问题有时被称作题项(items)。问题可以是提供答案的选择题,也可以是开放回答的问答题。问题和回答的选择、措辞和排列都需要仔细考虑,并注重语言技巧。

■ 示例 **选 题**

与问答题相比,多重选择题的主要优点是什么?

□ 可以客观地评分

□ 适合测量复杂的行为

□ 能够有不止一个答案

□ 对问题类型的威胁最小

■

■ 示例 **开放式问题**

与问答题相比,多重选择题的主要优点是什么?

请在此处作答

____

____

____

____

■

调查样本和设计

调查是一种从人那里获取关于人的信息的一种数据收集方法。从什么人那里、关于人的什么、多高频率以及什么时候?当你提出上述这些问题时,你就已经在关注调查的样本和设计。样本是调查涉及的人的数量和特征。设计指的是隔多久开展一次调查(只进行一次或截面的,还是多次的或历时的),参与者是随机选择的还是通过其他方式选择的,以及包含多少个不同的群组。

参考下面三个调查:

<u>调查1:2012级毕业生对健身了解多少?</u>

调查方法:在线问卷

样本:所有的1 000名州立大学2012级毕业生

多久进行一次调查:只一次——在毕业时

如何选择参与者:所有毕业生都符合条件

多少组:只有 1 组——2012 级

设计:截面的

调查 2:2012 级毕业生的健身知识在 12 个月内是否发生了改变?

调查方法:在线问卷

样本:所有的 1 000 名州立大学 2012 级毕业生

多久进行一次调查:两次——一次在毕业时,一次在 12 个月后

如何选择参与者:所有毕业生都符合条件

多少组:只有 1 组——2012 级

设计:历时性群组

调查 3:随着时间的推移,毕业班的健身知识是否有差异?

调查方法:在线问卷

样本:从 2012、2014 和 2016 级毕业班中随机抽取一个 75% 的样本,共 2 250 名毕业生。

多久进行一次调查:三次——一次在毕业时,一次在 12 个月后,还有一次在 24 个月后

如何选择参与者:随机

多少组:3 组——2012、2014 和 2016 级

设计:历时性和比较性的

调查 1 了解 2012 级毕业生健身知识的概貌,采用了邮寄问卷。这种概貌被称作截面调查设计。调查 2 想知道 1 年时间内(从毕业到之后的 12 个月)健身知识的变化情况。这个设计是历时性的。整个年级构成一个同期群。

调查 3 是历时性的,因为调查数据是从三个时间点(毕业时以及毕业后的 1 年和 2 年)毕业的三个毕业班中分别收集的。该设计还是比较性的,因为可以对三个毕业班在毕业时、一年后、两年后或所有三个时点的知识进行比较(或两两进行比较)。调查 3 的设计说明可以采取这种形式:

	何时进行调查		
年 级	毕业时	毕业后 1 年	毕业后 2 年
2012			
2014			
2016			

调查 3 与调查 1 和调查 2 的不同还在于毕业生如何被选中参与调查。在调查 3 中,一个 75%的毕业生样本被随机选中参与调查。在其他两个调查中,所有的毕业生,而不仅是一个样本,都可以入选。随机选择意味着每个毕业生拥有同等的机会被选中。

所有的三个调查都是在线的,但是它们的样本和设计各不相同。

数据分析计划

不管你的调查设计或规模是怎样的,你都必须事先考虑如何分析调查数据。

- 你打算计算百分比产生类似下面的结果吗?

在全部样本中,50%的人报告他们是共和党人,42%是民主党人,5%是独立派,1%属于绿党,3%是无党派。

- 或者像这样计算平均数?

回答者的平均年龄为 56.4 岁。教育水平的中位值是 13 年。

- 在组间进行比较而做出如下报告?

在男性当中总共有 60%的人是共和党人,女性当中只有 20%。

回答者在对现政府的满意度上没有显著差异。

- 你会寻找诸如此类的关系吗?

调查发现人们的自由或保守程度与其教育成就之间没有关联。

30 岁及以上的高中毕业者比 30 岁以上受教育程度较低的被访者显著地更加有可能在最近一次选举中投票。

- 你会关注历时性变化吗?

自从 2007 年以后,每天参与 2 小时或 2 小时以上托儿的人数在统计上差异显著。

小规模测试

小规模测试是尝试性的,目的在于形成可用的调查表,以帮助获取所需的信息。所有调查在付诸实施前都应进行小规模测试。自填问卷高度依赖语言的清晰性(至于其是书面的还是在线的则不要紧),而小规模测试能快速显示人们是否理解你提供的指导语,以及他们能否回答调查问题。面对面访谈的小规模测试还能告诉你访谈者的一些情况。他们能否轻松跟上访谈表?书面调查预留的空白够不够写录答案?在计算机辅助访谈中,访谈者是否知道如何处置“死机”的情况?在线调查中被访者是否知道如何前进和后退?小规模测试还可以告诉你完成调查需要花费多少时间。

测试可使调查更为流畅。只要有可能,就应该尝试复制调查发生的环境。这可能意味着需要获取人们只参加测试而不是整个调查的许可,虽然他们有资格参与整个调查。

回答率

调查者希望每个合格的人都回答所有问题。小规模测试有助于提高回答率,因为它能消除带来困难的潜在来源,比如措辞不当以及书面问卷中没有记录回答的空白。此外,如果整个调查程序都经过仔细测试,那么这也能帮助提高回答率。在开展电话访谈前,提出下列问题:你是否有最新的电话号码表?你是否希望在能够找到调查被访者的时候致电?是否计划好如何接触那些不回应电话留言或语音邮件的被访者?对于在线调查提出下列问题:是否有最新的电子邮件地址表?你是否知道如何确保回答的隐私?还有一些方式对所有调查类型,都有助于提高回答率,如使调查简短和简单,以及对参与调查提供奖励。

回答率多高合适?如果你正在进行一项大型的复杂调查,你可能希望通过统计程序回答这个问题。如果你的调查相对简单(比如 1 所学校的一组教师或 3 所医院的护士),那么你需要确定需要多少人才能使结果可信。如果有 20 个人符合参加一个邮件自填问卷的调查,

而只有 10 个人回答，你的感觉可能会与另外一个 400 人中有 200 人回答的感觉不同。每个调查都是 50%的回答率，但基于 20 人中的 10 人的报告看上去不如基于 400 人中的 200 人的说法那么令人信服。除非从统计上加以确定，否则期望的回答率完全是主观的，一般的原则是“越高越好”。

报告结果

互联网、电视、报纸和杂志上每天都在报告调查结果。对许多人来说，调查通常是某些人（而非所有人）对当前政治、社会或经济问题的关注情况。调查结果一般是像下面这样的。

■ 示例　　**调查结果的样子**

问题：如果今天举行选举的话，你是否会投票给候选人 X？

	是	否
男性	50%	50%
女性	20%	80%

■

■ 示例　　**调查结果的样子**

What is your favorite color?

Frequency Analysis

Answer	Count	Percent
Red	1	50.00%
Blue	0	0.00%
Orange	1	50.00%
Green	0	0.00%
Total	2	100%

Key Analytics

Mean	2.000
Confidence Interval @ 95%	[0.040 - 3.960] n = 2
Standard Deviation	1.414
Standard Error	1.000

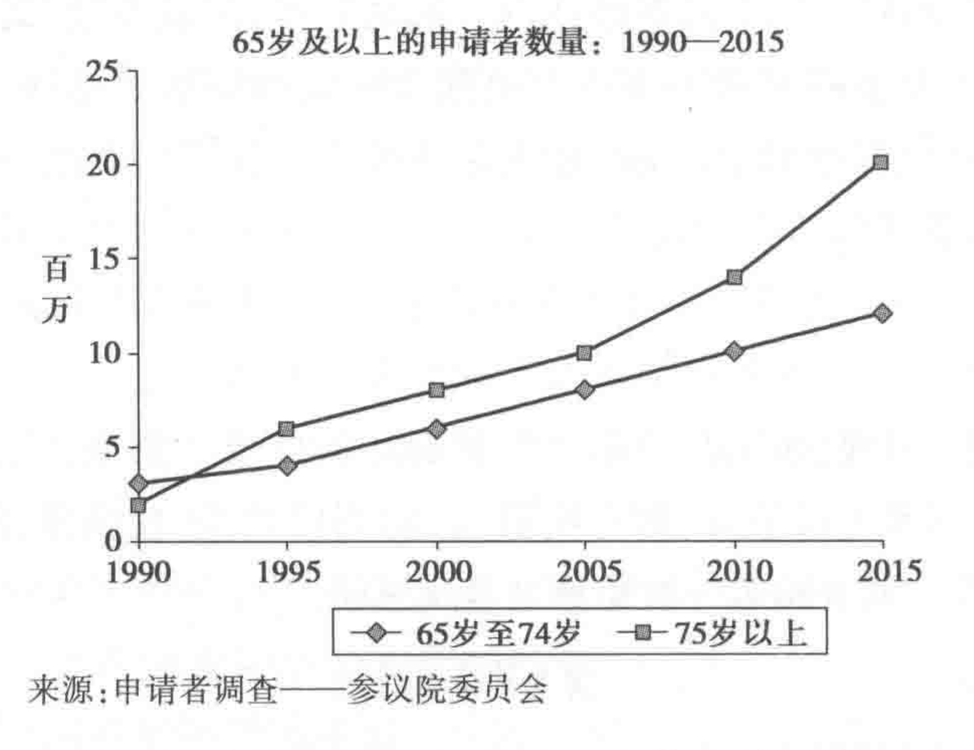
Key Facts

100% chose the following options :
- Red
- Orange

来源：选自 www.questionpro.com

■

■ 示例　　调查结果的样子



来源:申请者调查——参议院委员会 ■

关于上面例子中出现的标准误的更多信息参见第 5 章,置信区间和标准差的更多信息参见第 6 章。

所有调查,为了得到结果,都遵循以下必需步骤:

- 确定调查类型(邮寄或在线问卷;电话或面对面访谈)
- 选定调查的内容以及撰写问题和测试调查表
- 确定谁应该参与调查(所有人还是一个样本?)以及参与频度(只一次还是 5 年中每年 1 次?)
- 执行调查(谁来开展访谈? 什么时候必须提交在线调查表?)
- 处理数据(如何录入数据? 手工还是自动将调查数据录入数据库?)
- 分析和解释结果(回答率是否足够高? 人数多少或人们的差异意味着什么? 人们对候选人 X 的感觉如何? 意见是否随时间而改变?)
- 口头报告结果或采用文字、图表和图形的书面报告结果(谁是受众? 报告需要多长? 需要呈现哪些数据?)

任何可信的调查都不能忽略其中任何一个步骤,不过,由于目的和资源的不同,某些调查可能会比其他调查更重视某些步骤。

各种调查方式的利弊

你如何在自填问卷和访谈之间做出选择?邮寄与在线调查哪个更好?什么时候访谈是最佳选择?以下是一些选择不同调查类型的标准。

信度和效度

可信的调查带来一致的信息。有效的调查带来准确的信息。获得可信且有效的调查必须确保你用来选择问题的定义和模型是基于理论或经验的。没有哪种调查类型在一开始就有比其他类型更高的信度和效度。必须针对你的具体目的选择最为精确和准确的调查方法。例如,如果你担心被调查者不能很好地阅读,那么口头(面对面或电话)访谈可能会比书面形式产生更好的结果。

焦点小组和先行测试有助于你选择调查类型,确定调查设计是否是良好的、是用户友好的。回答者或调查执行者(那些访谈或发放问卷的人)如果对调查有疑问就会做错事,带来偏差,这会降低调查的准确性。一个设计良好、易于使用的调查对信度和效度有很大助益。

结果的有用性和可信性

如果结果是有效的,并且调查方法被用户认为是正确的,那么结果将是有用的。在开始调查之前就要找出人们期望的方法。有时候使用调查结果的人具有很强的偏好。

成本

这是指开发和执行每种调查类型的财务负担。成本包括书面问卷(现场或邮寄)的纸张费、复印费和奖金。邮寄问卷需要最新的通讯录(可能不得不购买)、邮资和信封。有时候还需要几封后续邮件,这也增加了成本。

面对面和电话访谈的成本包括电话费,以及大型调查中购买电话

系统的成本,还有为误拨、过时的电话号码以及挂机所付出的费用。你还需要支付为访谈者撰写脚本、培训访谈者和对被访者提供奖励的费用。

在线调查需要密集的开发和测试。节省开发费用的一个方式是购买用户友好的在线商业调查程序(像 Questionpro.com,还有其他程序)。但这样,你将不得不去学习开发软件,而这会是非常耗时且昂贵的。并且即使是最好的软件也不能弥补草率的调查产生的失误。

在线调查的特殊情况

在线调查是自填问卷。被访者在笔记本、台式机、平板电脑甚至手机上完成。调查者喜欢在线调查,而被调查者正在逐渐适应它。调查者喜欢在线调查是因为它能很容易地接触到世界范围内的大量人,还因为在线调查软件容易获取且相对便宜。

如果没有事先计划好,在线调查有时候就会遇到麻烦。调查通知可能被当做垃圾邮件删除,或是被发送给一个邮件列表中的所有人,而其中的很多地址是错误的。其他麻烦包括没有给答题辛苦的被访者提供足够奖励,或者问题太多或难于回答。和其他调查类型一样,调查者需要仔细决定问题和回答的内容和形式、样本量、谁应该谁不应该参与调查以及如何保证结果的准确性和安全性。设计和执行的不良无法依靠在线调查软件得到弥补不佳的调查。

下面的清单可以用来决定何时使用在线调查。

[决定在线调查最佳使用时机的清单]

- 调查需要覆盖广阔的地理范围。
- 你希望调查非常大量的人。
- 你可以获得有效的邮件地址并被允许使用它们。
- 你或你的机构或单位经常开展调查,因而任何在技术和软件设施上的投入都是值得的。开发软件去创建调查问题是十分昂贵的。有大量商业公司可以帮助设计和执行调查并分析数据。这部分很容易,然而总是存在一条学习曲线,学习

新软件需要花费时间,而时间需要财务资源。

- 你设计的一些调查问题所有被访者都回答了,但是另外一些问题只有部分被调查者回答。例如,如果你想探讨人们的锻炼习惯,但并非所有人都报告了他们的锻炼情况,你只想询问锻炼者的活动情况。在线调查可以通过程序分辨谁应该回答问题谁应该跳过。这使调查者不必使用过多复杂的指示语。
- 调查者希望使用多媒体(如视频片段和画外音)来描述一个新想法、产品或服务。
- 你需要尽快得到结果。在线调查可以获得每个人完成问卷的实时数据。注意,大多数调查软件包给出的统计结果都是相对简单的:数量、平均值,等等。图形也大多是简单的饼图和线图。更为复杂的分析和报告需要花费更多时间和专业知识。
- 你希望比较不同时间的调查数据。在线调查数据可以很容易地被存储在数据库里供将来使用。来自同一个或不同群体的新数据可以随时添加到数据库中,你可以针对一个或多个群体不同时间的数据进行比较。
- 你计划研究调查行为本身。对于在线调查,完成调查的日期和时间以及在每个问题上花费的时间都是已知的。这些信息可以用来判断人们在完成全部或部分调查上所需的时间是否存在差异。这些数据可以帮助你改进调查问题。

大多数网络调查都有至少 4 个潜在的注意事项。它们是:

1. **多地址**。许多人有很多电子邮件地址或使用社交媒体的在线联络点。虽然有些潜在的回答者会把他们的地址集成在一起,这样查收的时候就能看到所有来源的信息,但其他人并不会这样做。有些人经常只查收 1 或 2 个地方,还有些人不经常查收电子邮件。确保你拿到的地址是准确的和经常使用的,否则抽取出一个期望的样本或获得满意的回答率就会有问题。

2. **隐私和保密**。许多人对通过网络发送信息持怀疑态度。调查

者必须学会保护回答者的隐私,并确保让回答者感受到他们的努力。很重要的一点需要注意:不是每个人都知道他们的电脑或平板是否安全以及如何确保安全,他们也不知道谁曾经偷窥过他们的浏览历史。当然,如果被调查者设备的安全性不好,调查者也无法控制隐私的泄露。

3.**回答者的身份**。调查者没有什么实际的方式了解究竟是谁完成了问卷。

4.**技术差别**。回答者的网络接入、浏览器和字体颜色的设置各不相同。有些调查不允许使用浏览器的回退功能,而是设置了一个特殊的回退箭头。必须在所有常用的浏览器中对调查做测试。如果必要,需要指导回答者如何在调查中回退和前进。如果计算机(台式机、笔记本、平板电脑)掉线,一个好方式是设置一种机制,确保在停电或掉网的情况下回答者能够回到原来的位置。

许多网络调查是由出售软件和调查处理服务的卖家完成的。许多公司可以为调查者操办一切事务。他们可以发出调查、追踪回答者、分析数据和撰写报告。他们通常不会发起调查,但可以提供中介服务。

如果你决定购买在线调查软件,得准备好花上一周时间学习如何使用。大多数卖家都会在如何选择问题类型、问题和页码的数量、背景色及图标的特征等方面提供良好指导。无论你将从调查供应商那里获得多少帮助,十分重要的一点都是仔细阅读软件使用细则。

调查者依赖供应商更新和维护调查及调查数据。但是服务合同总是有限的,因此必须遵守这些细则。

[与商业调查公司合作的调查者应遵守的细则]

- 经常下载你的数据,把文件存放在安全的地方。服务中断后你可能就拿不到你的数据了。
- 对每项在线调查都创建一个离线副本,这样你就能把它们转化成纸笔形式的,或者在另一个在线服务中使用它们。确保

可以在你的计算机中保存调查。

- 使用商业公司建议的调查问题列表要谨慎。它们是否适合你的被访者？有没有证据表明它们曾经被使用过？谁使用过它们？

你可以到网上寻找如何评估供应商的信息。可在任何的搜索引擎使用这样的关键词“比较网上调查(compare online surveys)”。

移动电话调查的特殊情况

许多曾经依赖固定电话抽样的大型调查现在也开始使用移动电话样本了。不幸的是,调查者逐渐发现移动电话调查要比预想的昂贵。美国联邦法规禁止自动拨打手机,而手工拨打会增加成本。此外,筛选合格的被访者也会增加成本。现在,许多未成年人(17岁及以下)使用手机,但是年长的人则偏向固定电话。为了找到户主或者年长者这样的合格被访者可能需要付出额外的努力和花费。

移动电话调查也是复杂的,因为它们涉及的是个人而非地域,因此必须做出额外的筛选以确保被访者居住在调查者希望收集数据的地区。如果考虑到有些人有固定电话,有些人有固定电话和手机,有些人只有手机,这个过程就会更为复杂。没有人真正知道这些群体之间是否有差异,以及如果存在差异的话是否会影响调查的结果。

移动电话调查需要考虑被访者就直接或间接地为通话付费的情况。在美国,这类调查几乎总是要因此为参与调查者提供奖励。

做出决定

为了帮助你在相关的调查类型之间做出决定,表1.1比较了主要的调查类型的优缺点,并提示出它们特殊的要求和花费。

表 1.1　比较调查类型

	自填		访谈	
	邮寄	网上	电话	面访
特点	被访者完成的是书面或纸笔形式的调查。	调查是通过能上网的设备接触并完成的。	访谈人使用固定电话或移动电话调查被访者。	访谈人和被访者在访谈期间处于同一个地点。
优点	能够涉及广大的地理区域。许多人习惯纸笔形式的调查。人们可以在任何地方完成调查。	能够立即(实时地)获得世界范围内的信息。能够根据需要为被访者提供陌生词语的解释,以及帮助他或她回答困难问题。能够提供音像以阐明观点。易于向许多被访者发送提醒。易于处理数据,因为回答被自动地下载到电子表、数据分析程序或数据库中。不需繁复的指导语就可以根据被访者和调查者的需要对问题作出调整。	可以与被访者探讨答案。遇到不熟悉的词时可以为被访者提供帮助。有些人喜欢电话沟通。不用担心访谈人的容貌。	你可以与被访者探讨答案。有些人更愿意直接和访谈人打交道。你需要考虑访谈人的外貌(年龄?种族?着装规范?)
缺点	你需要激励样本完成一个复杂调查。 许多人认为他们太忙而无法完成调查。 被访者必须能读、看和写。	调查者需要可靠的电子邮件地址。 被访者必须有可靠的网络接入。 在不同的浏览器和屏幕上问卷的样子不完全一样。 一般的电子邮件地址无法选择随机样本。 可能遇到电脑死机或掉线的情况。	越来越多的人以手机替换了固定电话;还没有办法对只有手机或宁愿通过手机接受调查的人抽样。 你需要培训访谈人。 你需要确保能够找到被访者,有办法应付答录机和语音邮件的信息。 你需要确保手机被访者处于一个可以专心接受调查的地方(如,不是在驾驶中)。	你需要培训访谈人。 你必须找到一个合适的地方开展访谈,要想符合被访者的日程安排不一定很容易。

特殊要求	能够拿到最新的邮件地址列表是基本条件。催促邮件几乎总是必须的。奖励能增加回答数量。数据录入和分析不像网上调查那样是自动的。	最新邮件地址。与网上供应商接触所需的技术专业知识和资源。确保私密性的可靠方法。出现在被访者电子邮件程序“来自”一栏的名头应该响当当的。奖励。	最新电话号码。针对有固话和手机以及只有手机的人的精妙的抽样方法。被访者的年龄和其他重要个人特征可以差别大。奖励。	如果是在现场，那么需要空间和隐私性。进入某人的家可能是困难的。奖励。
花费	反复寄信所需的印刷、纸张、信封、邮票，对被访者的奖励。数据录入、编程和分析。	如果你准备设计你自己的调查，那么就需要大量的编程时间。如果你使用网上服务，就必须付出许可费，并且学习如何使用公司的软件。给被访者的奖励。	培训和监督费用、酬金、电话和电话费、计算机和技术专业知识费用以及给被访者的奖励。	培训、场地、差旅、给被访者的奖励。

调查连续谱:从具体到一般的应用

调查正在成为一种收集数据以回答关于健康和社会、经济和政治生活方面问题的主要方法。一项调查的广泛性和科学性应该多高才合适?

比较下面两个调查。

■ 示例 **具体用途的调查**

"邻里驿站"的主任希望向居民提供适当服务。当前,人们对于缺乏足够的健身设施有许多抱怨。计划开展一项对 5 家保健提供机构、100 位居民和 10 名全职或兼职雇员的调查,以找出需要以及能够提供哪些设施。

■

■ 示例 **一般用途的调查**

郡卫生部门关注其 10 所驿站的效益。10 个站点统共有 20 000 名居民和 220 名全职或兼职雇员。郡卫生部门与许多公共或私营部门的提供商协商保健服务的安排。作为其效益研究的一部分,郡卫生部门调查了一个来自所有站点的居民、雇员和提供商的随机样本。NextDoor 在线社区对于采用该郡卫生部门的驿站模式很感兴趣,热切期待调查和评估的结果。

■

第一个调查的理由是基于一家驿站对其自身需求的关注。第二个调查的理由则是基于郡卫生部门对其所有驿站效益的关心。另外,NextDoor 在线社区也对调查的结果感兴趣。调查 1,因为其有限的影响,可以在方法上不那么正式。而调查 2 则必须在抽样计划、问卷构造和数据分析及解释上保持严谨。调查 1 关注的主要是有用性。

调查 2 还关心效度和概化程度:如果换一个地方(NextDoor 在线社区),郡卫生部门的驿站模式是否还同样有效?

每次你做一个调查,都要评估其目标落在从具体到一般的连续统中的哪个位置上。你必须给满足某些具体需求的调查留有余地,而所有旨在推广研究发现的调查则都必须严谨执行。

伦理、隐私和保密性

有些人对调查持怀疑的态度。他们担心所提供的信息被不当使用。有许多技术可以保护个人的隐私,确保信息只在知情的情况下且目的明确时被使用。调查者应该使被调查者确信每项调查都已经采用了这些技术,所做的是道德上正确的事情。

所有完成的印刷或书面调查都应该封存起来,只有少量员工可以基于"必须"接触它们。在得出数据分析的结论后,应该把调查销毁。此外,你可以为每个人分配一个号码,用号码将人与其回答关联起来,这样就能把识别信息(如姓名、生日、社会安全码或政府颁发的身份号码)与调查回答分隔开来。网络调查的接受者在登录接受调查时可以被允许使用自己的账号和密码。

调查的应用和对伦理问题的关注是完全交织在一起的。开展调查的目的是了解情况;伦理考虑则保护个人的隐私权甚或匿名性。

如果你的调查是服务于接受美国政府资助的公共或私立机构的,那么你应该知道联邦政府对知情同意、隐私和保密有特别的法律规定。这些规定包括:

- 充分解释每个步骤及其目的
- 对任何风险和利益做出描述
- 对所有质询做出解答
- 指导人们如何收回同意以及不受损害地退出参与

保密性由"联邦法规汇编"(有时也称作"共同规则")的"人类被试保护"原则加以保障。保密性指的是保证关于某人的任何信息不被其他人所知。一个拥有人们姓名和地址——即使是以编码的形式或者没有身份信息("去身份化的")——的调查者,不可以使用这些信息去透露身份。在许多调查中,保密性是最受关注的,因为完全的匿名化实际上是不可能的。一个号码、电子邮件地址,或者有时候仅仅是一个邮政编码,都能导致被调查者的身份泄露。

如果你为私人机构、组织或公司工作,你应该遵守有关知情同意

和保密的规则。是否存在你必须事先获得其允许的人类被试保护委员会或机构审查委员会?如果你是一名学生,要核对一下你是否可以提出你想问的问题。还可能你属于某个较大的项目,该项目已经取得行动许可,只要符合某些标准——其中就包括被访者的知情同意。

知情同意

同意书要给潜在的被访者提供充分的书面信息,以便他们决定是否完成调查。这里是知情同意书中应该出现的事项。

[知情同意书的内容清单]

- 一个诸如"同意参加调查"的标题
- 调查的名称
- 调查的目的
- 开展的步骤,包括调查在何地举行及持续时间
- 可能的风险和不适,包括回答私人问题或者呆在封闭的室内2小时之类。
- 对被访者和社会可能的好处,包括关于开发中项目和制定中政策的新知识或有利信息。有时候好处并没有被真正意识到。
- 向参与者支付费用。告诉参与者能得到多少钱;如果没有报酬,要如实相告。
- 保密性。如果对被访者的姓名保密,要描述编码程序,谁可以接触到调查,完成的调查保存在哪里。如果信息会与谁共享,要说明到底是谁。法律可能要求你透露调查结果。
- 参与和退出。参与者可以随时退出吗?如果他们退出会发生什么?(例如,他们是否仍能获得奖励?他们是否仍能获得期望的教育、社会利益或健康保健?)
- 调查者的身份。如果对调查有疑问应该给谁打电话?

互联网与符合伦理的调查

网上调查涉及彼此相连的能上网的设备。在调查者和被调查者、被调查者和网络服务器以及网络服务器和调查者之间会发生通讯。如果不做适当的保护,网上随处都可能出现安全漏洞。

• 调查者和被调查者之间的通信

调查者通过电子邮件与被调查者联系很常见。电子邮件可能会讨论调查并通过点击一个网址或将该网址贴在诸如 Mozilla 的 Firefox、谷歌的 Chrome、苹果的 Safari、微软的 IE 等浏览器中邀请被调查者参与调查。不幸的是,电子邮件不总是安全或私密的。许多人不知道他们的计算机是否安全,甚或不知道如何保证安全。由人工维护的电子邮件程序经常不是私密的。如果人们忘了登出或者疏于保管密码,他们的隐私保护措施很容易遭到破坏。另外,弱密码也很容易被攻破。如果你开展一项调查要求人们使用密码,必须确保密码的设置是安全的。

• 被调查者和网站的通信

当一个被调查者在网上问卷的空白区里输入敏感信息时,比如购物者在网上购物时提供信用卡号。网商使用 SSL 协议保证互联网通信的安全。SSL 协议将使用者的调查输入加密(转化为编码),信息到达网站后再“解密”。许多被调查者逐渐意识到如果没有保障他们的回答很容易被截获,因而如果不保证回答被加密,他们就会拒绝接受调查。你必须事先决定是否使用 SSL 协议,以及如何向你的被调查者解释安全方面的安排。

• 网站和研究者之间的通信

在传送过程中,必须使用 SSL 协议或者安全文件传输协议保护敏感的身份数据。

• 数据保护

有些人排斥接受网上调查,甚至因为担心隐私受损而不愿上调查

网站。不管是由商业公司还是个人创建和维护的,所有存储敏感信息或身份信息的数据库都必须加以保护。加密数据库可能提供了最大的安全性。

所有有声望的组织都会开发或采用确保被调查者的隐私得到尊重的规则。以下是隐私政策的一套最低标准:

[调查组织者的隐私政策的最低标准]

1.准确描述调查者的数据库中将会存储哪些调查数据。
2.解释为什么某项数据会被保存。
3.说明组织是否会提供、出售或转交信息,如果是的话,是向谁以及在何种情况下。
4.告知网站如何监控对网站内容未经授权的改变。
5.讨论由谁来维护网站。
6.如果有关的话,解释是如何使用 cookies 的。cookies 是你的浏览器存储的少量信息。cookies 允许网络程序存储关于项目选择、用户设置、登录信息以及其他一些可供今后调取的信息。cookies 会话是特殊的吗?如果不是,用户是否可以选择退出在会话之外 cookies 中存储的网页特征。

以下是疾病控制中心的隐私政策(http://cdc.gov/privacy.html)。虽然该中心的网站有多种用途(不是只为调查或调查报告),但它的隐私政策易读,且说明了如何满足隐私政策的最低标准。

■ 示例 **疾病控制中心隐私政策的介绍**

感谢您访问疾病控制和预防中心(CDC)的网站并浏览我们的隐私政策。除非您同意提供个人信息,否则疾控中心不会在您访问我们的网站时收集这些信息。疾控中心的隐私政策如下所述:

在访问疾控中心网站期间,如果只是简单地浏览网站或下载信息,关于本次访问的某些信息会被自动地收集和存储下来。这些信息不会暴露您的个人身份。我们只是自动收集和存储以下信息:[**对存储在 CDC 数据库中的信息的描述**]

• 网络域名(例如:使用个人网络登录账号时的"xcompany.com",或者

通过某大学域名连接时的"schoolname.edu")和IP地址(IP地址是上网时自动分配给电脑的一个数字);

- 操作系统以及登录网站使用的浏览器信息;
- 访问日期和时间;
- 是否将www.cdc.gov设定为主页;
- 访问的网页;以及
- 在访问CDC网站之前刚刚访问的网站的网址。

收集和分析这些汇总信息有助于我们提高网站的表现力,改进我们网站提供的信息材料的质量,以及提升总体的客户服务。如果向CDC发出了邮件信息或者完成了CDC网站上的问卷、表单或其他在线调查,CDC将在符合联邦法律的前提下保存这些信息。[**为什么数据会被存储**]

CDC不会泄露、出让、售卖或者转移任何关于CDC网站访问者的信息,除非应执法人员要求或符合相关法律。[**数据是否会被出让、售卖或转移**]

出于网站安全的目的以及确保为所有使用者提供这些服务,CDC采用软件程序识别未经授权的上传或改变信息的尝试,以防引发危害。[**在监控未经授权的使用方面是怎样做的**]

本网站由美国政府维护,受到美国法典第18条各条款的保护。违反第18条将被法庭刑事起诉。

某些CDC网页使用了"cookies",这是存储在你的计算机的浏览器中的一小段信息。

Cookies的使用

CDC网页使用两种类型的"cookies":基于会话的和持续的。[**使用的cookies类型**]使用的大多数cookies是基于会话的。[**cookies是如何使用的**]这些少量信息以一行文字的形式临时存储在你的计算机的随机存取存储器(RAM)中。当你关闭浏览器,cookies就会被销毁,不会在你的计算机硬盘上留下任何痕迹。使用这些基于会话的cookies只是为了提升访问网站者的用户体验。在很少的情况中,CDC网站使用持久的cookies较长时间地存储信息。在这些情况中,网页都会有明确提示,使用者也可以选择不使用这些网页特征。[**使用者可以选择退出**]如上所述,CDC在任何情况下都不会泄露、出让、售卖或转移关于CDC网站访问者的任何信息,除非应执法人员要求或其他法律要求。■

再看一下下面这个节选的隐私声明。这个声明来自一个开展调

查的非常大的公司。正如你所看到的,这家公司对于其他公司追踪顾客行为的可能性是诚实的。不过,有义务让顾客知晓:(1)意识到某些不恰当的 cookies 可能被植入他们的硬盘以及(2)如果他们愿意,可以联系隐私管理人员解决问题。

■ 示例　　**节选自一家非常大的公司的隐私声明**

我们的某些业务伙伴会在我们的网站中使用 cookies(例如,指向业务伙伴的链接)。我们不希望我们的业务伙伴在顾客离开我们的网站后使用 cookies 追踪其行为。不过,我们对于业务伙伴在我们的网站上如何使用 cookies **并没有完全的控制**(黑体为本书作者添加)。如果你觉得有某个[公司名]业务伙伴在您的硬盘中植入了不恰当的 cookies,请您与我们的隐私管理人员联系以帮助我们解决这一问题。 ■

这个节选的隐私声明提出了几个问题:当硬盘中有 cookies 时被访者是否确实能知道?被访者如何才能找到隐私管理人员?相关信息可以在网站的"联系我们"部分获得,但被访者不得不去查找。这些问题是有意义的,因为公众对在线调查的疑虑在增长,希望知道他们的数据是如何被使用的。

如果你能确保在向网上发送调查或调查信息时已考虑到这些隐患,就能帮助避免部分问题的发生。

如果你准备使用互联网(包括电子邮件)(1)与研究参与者联系或者(2)将参与者的信息发送给合作者或合同方,则你应该能够完成下列保障在线调查合乎伦理道德的问卷。

[保障在线调查符合伦理道德的示例问卷]

1.描述为了确保托管网站的网络服务器受到保护而准备采取的措施。在描述中提供关于物理安全、防火墙、软件补丁/更新和入侵演习的信息。

2.如果在登录网站时使用了密码或其他安全认证措施,需提出以下两个问题:

- 如何发送用户密码?
- 如何终止密码和网络登录?

3.如果对用户会话进行了加密,描述即将使用的加密方法。

4.解释何人拥有对服务器上数据的管理员权限。给出姓名、研究角色和所属单位。

5.详细解释限制非法和非必需登录的保障措施。

6.描述信息将被如何使用。是否会把信息出让、售卖或转交给任何人?

7.给出程序拥有者(也就是维护程序的人或人们)的姓名和地址。

8.如果使用电子邮件联系被访者,描述采取了哪些措施确保与被访者联系的是一个得到授权的人。

9.如果要求被访者通过电子邮件与调查者联系,描述如何保障是被访者本人发出的电子邮件。

10.解释研究知情同意书如何描述使用电子邮件对隐私的可能危害。

11.如果使用电子邮箱向调查者、供应商或其他人发送研究数据,解释电子邮件是否以及如何被加密。

12.如果被访者通过电子邮件发送附件给你,告诉他们附件是否会被加密或有密码保护。

13.如果使用了自动邮件路由系统,描述将会采取的安全控制。特别是描述测试和灾难恢复步骤。

14.如果合同方或供应商能够接触到调查被访者的身份信息或私密信息。

- 描述合同中保护被访者隐私所使用的语言;
- 描述对合同方或供应商(他们设计或托管着本项目的网络服务)的安全要求。

15.给出安全负责人的姓名,他负责调查组织关于本项目隐私和安全保护的政策和程序是否得到遵守。提供他或她的专业职务和单位。

16.给出具体负责本调查的与信息技术相关的一般安全保

障人员的姓名,提供他或她的专业职务和单位。

在需要收集什么和向谁收集方面,每个调查都有不同的局限。有些调查样本可能较其他人更为“脆弱”,因而需要不同的保护。下面是一个用于大学区教师网络调查的典型的知情同意书。调查者的目的是识别工作场所改进的需求。

■ 示例　　**一份网络调查的知情同意书**

本调查项目组及其调查合同方网络系统公司(Online Systems, Inc.)将会为你对每个调查问题的回答保密。需要保密的数据是诸如个人或学校的身份等信息。除非得到被访者的允许,否则这些信息不会泄露到调查项目组以外。个人可以授权项目组发布描述了他们自身的保密信息。调查项目成员学校的法定代表可以授权项目组发布描述了他或她所属学校的保密信息。[**评论:这里界定和描述了保密的有限性**。]

网络系统公司将提交包含学校和部门级信息的汇总报告,以帮助你的学校识别、优化和实施工作场所的改进以提高学生入学率。不会报告5人以下的被访者群体的信息。[**评论:有可能在非常小的群体里识别出个人的观点,而这会危及隐私性**。]开放问题的数据将在去身份化和重新编辑后提交给你的学校。项目组只会保留去身份化的个体层面的数据,只有去身份化的汇总分析才会在学术领域的出版物和研究报告中分享。[**数据将会被如何使用**]调查项目组可以把去身份化的回答提供给个人,只要他同意保护数据或者同意遵守项目组的保密政策。网络系统公司将在安全的服务器上存储数据,并将在调查完成后的两年内销毁所有的身份化信息。通过参与本调查,你将为你的学校贡献有价值的信息。[**评论:服务器将会是安全的。供应商必须在2年内销毁身份化数据**。]项目组和网络系统公司采取了大量步骤保护调查项目的参与者。伦理委员会的要求需要您知晓如果所收集的包含个人身份的信息被公开可能带来不良影响。[**这里是参与的风险**。]

根据调查项目组的政策和程序,这项调查已经得到审查和批准。继续之前,需要您承认已经阅读并理解了上述信息并同意参与本次调查。[**评论:这是一个网络调查,所以不要求被访者“签字”表明参与意愿。虽然有签字软件,但是绝大多数调查把完成调查视为知情同意**。]如果还有关于调查的任何问题,请联系……如还有任何关于你作为一个研究参与者权利的问

题,请联系……[**评论:有问题时联系谁**] ■

有些大的机构和公司有伦理委员会和隐私管理人员可以提供帮助使你的调查符合伦理要求。不过,很多调查者和调查公司并不熟悉有关隐私性的技术,在网络调查的伦理上也缺乏训练。为了多了解相关内容,你可以上网浏览国家健康研究所关于符合伦理的研究的指导意见(http://grants.nih.gov/grants/policy/hs/ethical_guidelines.htm)。机构合作培训计划(Collaborative Institutional Training Initiative)也提供人类被试的研究符合伦理方面的培训(http://www.citiprogram.org/Default.asp)。

小 结

调查是一种信息收集方法,用来描述、比较或解释个人性和社会性的知识、感受、价值、偏好和行为。

当你需要直接来自人的,关于他们相信、知道和思考什么的信息时,调查是最佳的方式。

调查可以是某人独自填写或在帮助下完成的自填问卷。自填问卷可以采用书面或网上调查的形式。书面调查可以通过邮件或在现场进行。

调查可以是面对面或电话访谈。

所有调查都由包含一系列问题,问题中人们有机会展示反应差异。这也是为什么问卷这个词经常与调查这个词互换使用。

为了获得准确的调查数据,你必须考虑调查的题项、回答选项、抽样方法、回答率、设计和数据分析。

调查结果以书面(印在纸上或电子复制)或口头报告的形式呈现。

可靠的调查会带来可靠的信息,而不可靠的调查会产生偏误的信息。

邮件调查因为人们熟悉它们而经常被使用。然而,如果不进行大量的跟踪和激励,回答率经常是不佳的。

访谈员需要进行培训,而这可能是耗时且昂贵的。

互联网是接触广大受众的一种有效方法,但无论是你自己做还是

有商业公司协助,专业知识都是必需的。像所有的调查一样,网络调查也需要事先周密计划。

手机调查可能是昂贵的,而且对于那些只有固定电话、同时拥有固定电话和手机以及只有手机的人,很难从中抽取样本。

可以出于具体或一般的目的开展调查。准备用于许多人和地方的调查结论,需要在设计方面给予特别关注。

所有已经完成的书面调查需要保存在上锁的文件柜中,而且应该只有少数人在需要的时候才可以接触到它们。在数据分析完成后,应该把调查资料销毁。

给每个人一个编码,以把各人的身份信息(如姓名、生日、社会安全号码)与其调查回答分离开来,而又保持人们与其回答之间的关联。

应该允许网络调查者在调查时自由选择自己的标识名和密码。

如果使用电子邮件地址通知参与者有关调查的事情并给出调查链接,那么这些地址应该只能被选定的某些调查者接触到。

如果公共或私人机构的某项调查得到了美国政府的资助,那么关于知情同意、隐私和保密方面有严格的法律要求,这些要求包括:

- 客观解释遵循的程序及其目的
- 描述任何可能的风险和收益
- 提供回应质询的方式
- 告知人们可以在不受歧视的情况下自愿收回同意以及停止参与

参与调查的知情同意指的是被调查者在明白下列事项之后仍然愿意完成调查的意愿,这些事项包括:调查的目的和程序,作答带来的可能的风险和收益,是否对参与调查付费,如何处理保密性,以及是否可以在不遭受损失的情况下退出调查。

一项调查的隐私政策包括以下内容:

- 描述将会被存储在调查者数据库中的数据
- 解释为什么这些数据会被存储
- 解释组织是否会出让、售卖或转交信息,以及如果是的话,是在何种情况下以及向谁

• 告知如何监控改变网站内容的非法尝试

• 讨论由谁来维护网站,以及如果有关的话,解释如何使用cookies。它们是基于特定会话的吗?如果不是,用户可否在会话结束后不允许网页保存cookies?

思考一下

阅读以下关于每个调查的描述,然后回答这些问题:

1.调查是关于什么的?

2.说明调查方法

3.为了达成调查目的你是否也会选择此种方法?请解释一下。

调查描述

通过“9·11”对美恐怖袭击的应激反应预测心脏问题

健康研究者希望考察2001年9月11日针对美国的恐怖袭击所带来的应激反应大小是否可预测其后3年全国性的心脏问题。他们在一个关于应激反应的网络调查中招募了一个全国性的成年人样本。参加调查的人在袭击发生之前已经完成了一个健康调查,因而研究者可以获得关于他们的“基线”信息。在袭击发生2周内,参与者完成网络调查,并在随后的第1、2和3年连续接受调查。研究者联系调查参与者的医生了解他们在3年中是否曾被诊断患有心血管(心脏)疾病。研究者发现对9·11的急性应激反应与随后3年中心血管疾病发病率上升53%有关。

学校家具与腰痛

这项调查的目的是了解某些类型的学校家具是否能防止或导致孩子腰疼。546名年龄在14~17岁的学生完成了一项书面调查,回答在校期间的坐姿以及腰疼的发生和严重程度。孩子的书包尺寸和重量,以及学校家具的类型和尺寸均被测量。研究发现在之前的3个月中有超过一半的青少年经历了腰疼,大约四分之一的人报告因为腰疼而减少每天的活动或求医。没有发现腰疼与学校家具的类型或尺寸或者孩子身高体重相关,但发现与单肩书包重量之间

存在正相关关系。

男女对爵士乐即兴演奏的参与

对一所主要的中西部大学 60 英里范围内的中学(含初高中)、初级中学、高级中学、大学和社区参与爵士乐项目的 332 名学生进行了调查。调查在爵士乐队演奏期间进行,约需花费 10 分钟。由爵士乐队主管和/或学生教师负责对中学和高中学生的调查,让他们把调查问卷带回家请家长签字表明允许孩子参与调查。研究发现女性对学习爵士乐即兴演奏显著地更缺乏自信、更焦虑、自我效能(态度)更低。

对有偿个人援助服务的满意度

美国传统的公共家庭护理项目依赖公共或私人机构雇佣或辞退护理人员、计划和指导服务、监控服务质量、必要时约束护理人员、支付薪酬及相应的工资税。在机构导向模式中,客户可以表达对服务和人员的偏好但却无法直接加以控制。这个调查主要是比较华盛顿州医疗补助项目的受益人对顾客导向和机构导向的有偿个人援助服务的满意度。调查主要通过电话访谈进行,使用了一套计算机辅助电话访谈系统,同时也进行了少量面访。调查发现,在老年人当中,而不是较为年轻的残疾人士中,接受顾客导向服务的受益人比接受机构导向的人的满意度更高。

青少年司机的领照偏好和实际行为

这项调查采用概率抽样方法从一个全国性美国家庭代表性样本中抽选出 1 383 名年龄在 15~18 岁的被调查者。对这些青少年进行了一项网络调查以了解全国的领照率、对尽早领照的兴趣和推迟领照的原因。有些青少年来自只有手机的家庭,为他们提供了网络接入。调查发现,16 岁是个中位值,这个年龄的孩子中差不多一半尚处在学习阶段,一半已经领取了正式或有限制驾照。对于已经到了可以领取驾照年龄的青少年,没有车、费用、是否有家长、没有车也能出行以及是否忙于其他活动都是导致推迟领取驾照的理由。

对下述某调查组织的隐私政策的充分性做出评论。

信息安全

调查组织的员工理解用户隐私的需求,我们采取一套严格的程序保护您的信息。组织指定了一位隐私政策管理人员督查隐私方面的问题。接触用户数据被严格限制在懂得尊重用户因素的特定人里。这些员工也只能接触基于业务原因所需的信息。对谁接触过数据的记录进行保存和监控,以避免安全问题的发生。

文 献

Bowker, A., D' Angelo, N., Hicks, R., & Wells, K. (2011). Treatments for autism: Parental choices and perceptions of change. *Journal of Autism and Developmental Disorders*, *41* (10), 1373-1382.

Holman, E.A., Silver, R.C., Poulin, M., Andersen, J., Gil-Rivas, V., & McIntosh, D.N. (2008). Terrorism, acute stress, and cardiovascular health: A 3-year national study following the September 11th attacks. *Archives of General Psychiatry*, *65* (1), 73-80.

Skoffer, B. (2007). Low back pain in 15-to 16-year-old children in relation to school furniture and carrying of the school bag. *Spine*, *32* (24), E713-E717.

Wehr-Flowers, E. (2006). Differences between male and female students' confidence, anxiety, and attitude toward learning jazz improvisation. *Journal of Research in Music Education*, *54* (4), 337-349.

Wiener, J.M., Anderson, W.L., & Khatutsky, G. (2007). Are consumer-directed home care beneficiaries satisfied? Evidence from Washington State. *Gerontologist*, *47* (6), 763-774.

Williams, A. F. (2011). Teenagers' licensing decisions and their views of licensing policies: A national survey. *Traffic Injury Prevention*, *12* (4), 312-319. doi: 10. 1080/15389588.2011.572100.

2

调查问卷:问题、量表和格式



为了确定调查的内容,你首先必须界定所要测量的态度、信念、价值或理念。例如,恐惧是什么意思? 自由主义观点呢? 自我效能呢? 还要问:如果希望测量恐惧(或自由主义观点或自我效能),在一个有效的调查中需要提出哪些问题? 能够从调查中获取所需的信息吗? 是否有其他更好的数据收集方法?

对封闭问题的回答可以采取是或否、选单和评级等形式。选单给出了回答的多种选择(如,“勾选所有适用的选项”)。评级要求被访者给想法分出等级(如:“1=最高,10=最低”)或排序(如:“1=完全同意,2=同意,3=不同意,4=完全不同意”)。被访者也会被要求为一个回答赋值(如:“你多大年龄?”)。

对问题的回答产生所谓的定类(如:“是或否”)、定序(如:“大、更大、最大”)和定距或定比(如:“20、21、22 岁”)数据。调查者需要考虑调查产生的数据类型,以匹配适当的分析和解释研究发现的方法。

调查者经常对个别项目的回答感兴趣,如会投票给候选人 X 的人数或者 65—80 岁之间的妇女在三个月内看医生的

频度如何。但有时候,他们也会关注一组项目的得分,这组项目综合反映了被访者的观点、健康状况或感觉。

网络调查有许多好用的特性,如下拉菜单和基于对一个问题的回答创建下一个问题,仅仅点击就可以接触到大量的人。至于成本,要包含购买和学习使用新软件的费用和时间。

内容即信息

一旦你决定采用调查的方法去收集数据,就必须考虑其包含的内容和主题。

假设你要评估一家青年中心的进展,你的主要任务是评价项目的目标是否已经达成。比如说目标之一是通过提供教育、工作、财务帮助以及医疗和心理健康援助来提高青年的自尊。再假设你准备调查一个青年项目的参与者以了解其自尊状况。那么你如何确定应包含哪些调查内容呢?

为了选定调查的内容,你需要首先界定术语,弄清楚你需要什么,能够通过提问从人们的观点中获得什么。

界定术语

许多人类的态度和感觉,像自尊这样的,拥有多种定义。自尊是否意味着对自己感觉良好?如果是的话,感觉良好的意思又是什么?调查者可以通过下列方式回答诸如此类的问题:综述研究(在本例中,也就是关于自尊的文献)以了解像自尊这样的概念已知的内容和已有的理论,咨询专家,或者给出自己对概念的定义。使用自己的定义的问题是别人可能会怀疑其效度。当使用像自尊这样的理论概念时,也许最好的方式是采用公认的观点,如果可能的话,甚至是采用已经存在并验证过的量表。

当然,对于许多调查而言,你可能无须测量理论概念,但即便如此,你也需要界定自己的术语。假设你正在评估一个社区对健康服务

的需求。需求和健康服务这两个术语就需要定义,因为你可以根据所需服务的类型和特征(门诊? 医院? 家访?)以及方便程度(一天当中什么时间应该有医生)或者持续性(是否有必要总是由同一位医生出诊?)来界定它们。

选择你的信息需求或假设

假定两位调查者在他们对青年中心的评估研究中使用了相同的自尊定义。根据情况的不同,调查者 1 可能更关注与总幸福感相关的自尊,而调查者 2 可能更关注在学校或工作中表现出来的自尊感。显然由于方向的不同,调查者 1 和 2 会提出不同的问题。其结果是会产生不同种类的信息。调查者 1,由于只关注总体自尊,甚至可能不会涉及工作或学校,因此也不会提供这方面的信息。调查者 2,由于对学校和工作特别感兴趣,可能不会提供与人际关系有关的参与者自尊的信息。每个调查揭示的信息显然是不同的。

比如说你对青年中心参与者在参与项目 2 两年后其总体自尊是否提升感兴趣,并且根据某个已被接受的青少年人格理论界定了相关术语。为了确保获得所有你需要的信息,你需要提出下列问题:我希望和必须获取因而必须收集哪些信息? 记住,如果你不提出这样的问题,后面你就很难报告它!

以下是青年中心的评估者可以提出的一些典型问题:

- 在总体自尊感和参与者是男孩还是女孩之间是否存在一种关系?
- 参与者的自尊是否依进入项目的时间长短而不同?

这两个问题表明调查需要收集三个主题的数据:

- 总体自尊感
- 参与者的性别
- 参与项目的时间长短

如果这些主题中的任何一个被忽略了,调查者就不能回答那两个评估问题。毕竟,在本例中,调查只是为了评估。

确保你能够获得信息

在某些案例中，人们可能不愿意表明他们的意见。例如，青年中心的评估者可能发现许多年轻人不愿意回答关于他们的感觉的问题。而在另外一些案例中，可能的被调查者只是不能提供答案。比如你希望了解加入青年中心6个月的参与者在他们刚刚进入中心项目时对学校的态度，许多人可能已经忘记了。无法回答问题并不罕见，在准确预测全国选举中投票人的偏好时这就是一个主要问题。人们在选定一个候选人之前要花费很多时间，期间有些人的意见可能数次发生变化。这就是同一个人及不同时点的投票结果各不相同的原因。

如果你不能确信从调查中获得所需信息，那么就删掉这个主题并寻找其他数据源，比如观察或者查阅诸如日记和学校出勤记录或入院证明之类的记录。

不要去问你无法施加影响的信息

在社区对健康服务的需求调查中，要求人们对外伤中心的偏好做出评价是不恰当的，如果社区并不能提供此类服务的话。记住调查的内容会影响到被调查者的想法和期望。如果你不能够或不准备满足它，为什么要去点燃希望之火呢？

一旦你选定了内容，确定了调查的边界，下一个任务就是实际撰写问题。写下的问题要多于你计划使用的，因为很可能会删掉一些不恰当的问题。第一稿中经常会有一些所有人都给出相同回答或者没有人能够作答的问题。在确定问题的数量和顺序之前，你必须确保调查包含了所有被认为重要的内容。你可以像下面的青年中心调查者所使用的清单那样，也列一个清单。如你所见，这个调查不会涉及员工敏感性，但会关注服务的周到性、可及性和可得性。

■示例 **青年中心满意度调查的计划**

主　题	问题数量
1.员工敏感性	
平时顾问的数量	0
需要时是否能够找到顾问	0
预约职员的礼仪	0
2.服务可及性	
时间是否方便	IIII
公共交通	IIII I
3.对参与者需求的周到	
翻译帮助	IIII
预约的容易度	IIII I
等候时间	IIII I
4.所需服务的可得性	
医疗	IIII IIII
教育	IIII IIII I

■

撰写问题

开放问题和封闭问题

调查项目采取问题的形式。

■示例 **开放问题**

1.为你做预约的人的礼仪如何？

____

____

____

■

或者也可以用陈述句。

■ 示例　　　　封闭问题

圈出你同意还是不同意以下各项:

2.为你做预约的人是礼貌的。

圈选一项:

非常同意	1
同意	2
不同意	3
非常不同意	4

■

有时候调查项目是开放性的,意味着被调查者同意用自己的话回答问题。问题1是开放性的。在其他时候,调查项目强制被调查者从预先设置好的选项中做出选择,如同问题2那样。

绝大多数调查依赖多重选择问题,因为它们被证明更有效率因而也更可靠。它们的效率来自于更易于使用、评分和录入数据。同时,它们提供的统一的数据也强化了其可靠性;每个人的回答都基于同样的选项(如:"同意或不同意"或"经常或不经常")。

开放性问题能够启发解释人们为什么相信他们所做之事,但解读这些开放性问题十分困难,除非建立一个精确的编码系统,并且有经过训练的人把收集到的数据归类纳入体系中。

看一看对一个问题的这两种回答,该问题来自对一个小学教学项目参与者的调查。

■ 示例　　　　小学教学项目的开放问题

问题:该项目最有用的三个部分是什么?

____

____

____

回答:

被访者A

指导员的讲座

现场体验

教材

被访者 B

指导员

课堂教学

最有用的部分是项目提供的浓郁的学习氛围 ■

对被访者 A 和被访者 B 的回答进行比较并不容易。被访者 B 列出指导员是有用的。这是说指导员是个总体上有用的资源吗?它与被访者 A 认为指导员的讲座有用的观点如何相比?换句话说,被访者 A 和 B 是否给出了相同的回答?被访者 A 说教材是有用的。如果项目中只使用了 1 种教材,那么被访者 A 和 B 的回答是相同的。但是因为记录下来的两个回答是不容的,一些猜测或解释就是必要的。

被访者 A 和 B 都提到了对方没有提到的一些东西:现场体验和学习气氛。如果它们同样重要,就应该分别予以分析。但假如在调查使用者看来两者都不是特别重要,那么它们是否能归入"其他"类?通常设置"其他类"来纳入分析和解释起来十分复杂的困难回答。在你了解它之前,其他有可能是最大的一类。

虽然被调查者回答开放问题相对容易,但是分析和解释是复杂的。以下封闭问题被用来收集类似的信息,但却使回答易于解释。

■ 示例 小学教学项目的封闭问题

圈选一项					
		完全不满意	不满意	满意	完全满意
a.	教材,课堂教学	4	3	2	1
b.	指导员的相关知识	4	3	2	1
c.	讲座题目的实用性	4	3	2	1
d.	现场体验	4	3	2	1
e.	其他,请指明	4	3	2	1

■

做出决定:开放还是封闭问题

当你希望被调查者用自己的话表达意见并且有足够的兴趣和资源去解释发现时,就选择开放问题。开放问题相对容易编写。

封闭问题因相对容易评分、分析和解释而被选择。但封闭问题可能难于准备,因为要求所有被访者对其的理解方式一致,并且所有相关选择都包含在内,且为互斥的和符合实际的。

整理开放调查项目的回答:你是否感到满意?

调查的一种常见用途是了解人们对一种新产品、服务或项目是否满意。他们的意见为理解为什么新的想法或做事方式起或不起作用提供了重要的启发。

一组开放问题特别适合收集有关人们最喜欢(LB)和最不喜欢(LL)某种产品或服务的什么方面的满意度信息。

以下是 LB/LL 技术如何发挥作用的。

步骤 1:询问被访者的意见

让被访者列出什么是好的以及什么是坏的。通常要为回答的数量设定一个限量;例如:“关于本次会议,请列出至少 1 件至多 3 件你最喜欢的事情。”如果被调查者的回答不足 3 个,他们可以空着不写或写上“无。”如果回答多于 3 个,你可根据你需要的信息决定保留或舍弃多出的回答。

步骤 2:编码 LB/LL 数据

一旦你获得了所有的回答,下一步就是对其分类和编码。为此,你可根据回答情况或者以往类似项目的经验创建分类。

分类越精确越好,也就是尽量分得细一些,因为其后如果有必要,合并它们要比分离它们容易。

假设以下是被调查者对“最不喜欢本次会议的什么”这一问题的

典型回答:

有些人说得太多

指导员不总是在倾听

我的发言总是被打断

有太多的噪声和混乱

有些参会者被忽视了

指导员没有施加控制

我没有发言机会

只有史密斯和琼获得了发言机会

指导员似乎漠不关心

我的思路总是被打断

你可以像下面这样分类和编码这些回答:

■ 示例

LB/LL 的例子:回答分类

	编码
指导员没有倾听(忽视参会者;表现得漠不关心)	1
有些人垄断了讨论(一直在说;没有发言机会;只有史密斯和琼有发言机会)	2
混乱的环境(噪声太大;指导员未施加控制;思路总被打断)	3
现在将编码与回答相匹配:	

■

■ 示例

LB/LL 的例子:参会者的回答

	编码
参会者 A	
指导员不总是在倾听	1
我的思路总是被打断	3
我的发言总是被打断	2
参会者 B	

续表

	编码
指导员不总是在倾听	1
当会场嘈杂时指导员没有及时制止	3
指导员忽视了某些学生	3
参会者 C	
我没有得到发言机会	2

为了确保正确分类,你需要考虑其可靠性。它们是否足够清晰,因而至少两个评分者将会赋予某个给定回答以相同的编码。

步骤 3:LB/LL 数据

当你对可靠性满意后,下一步是计数每个编码的回答数量。

下面演示对于 10 个参会者如何开展这一步:

■ 示例

LB/LL 的例子:参会者的回答

	编 码			
参会者	1	2	3	合 计
A	1	1	1	3
B	1		2	3
C		2	1	3
D		1	2	3
E		3		3
F		2	1	3
G		2	1	3
H		2	1	3
I		3	2	5
J		1		1
	2	17	11	30

看一下每类的回答数量。10 名参会者总共列出了 30 项关于小组讨论他们最不喜欢的事情。30 项中的 17 项属于同一个类别:编码 2。因此调查者可以有根据地认为,基于数据,关于研讨会参会者最不喜欢的是有些人垄断了讨论而其他人没有发言机会。

接下来,数一下每个编码下的参会者的回答数量。例如,只有参会者 A 和 B 做出了编码为 1 的回答。

■ 示例 **LB/LL 的例子:参会者的回答模式**

编码	本编码下做出回答的参会者的数量	是哪些参会者
1	2	A,B
2	9	除了 B 以外的所有人
3	8	除了 E 和 J 以外的所有人

■

看一下每一类中作答的参会者数量。因为 10 个参会者中有 8 或 9 个人的回答属于相同的两类(编码 2 和 3),他们的意见可能代表了整个群体的意见。还可以比较有把握地说参会者也不喜欢研讨会期间弥漫的混乱氛围。他们抱怨噪声令其无法清晰思考,而指导者没有施加控制。

当参会者的意见一致的时候,回答的类型将会很少,许多人都会做出这样的回答。当回答者意见不一时,就会有许多不同类型的回答出现,每类回答的人数很少(少于 10%)。

当你有许多参与者和回答需要分类时,对 LB/LL 数据的解释会变得更为复杂。例如,假设你向 100 名参与者提问对某个健康教育项目他们最喜欢什么方面。

首先,你必须确定你的回答分类,并赋予每个分类一个编码。接下来:

1.按序排列编码。也就是说,如果最大数量的参与者选择的回答的编码是 3,那么就首先列出编码 3。

2.计算分属每一编码的学生的百分比。如果 100 名学生中有

40 名做出了属于编码 3 的回答,那么计算出来就是 40%。

3.计数每一编码的回答数量。

4.计算属于每一编码的回答的百分比。如果 400 个回答中有 117 个回答属于编码 3,那么编码 3 所占回答的比重就是 29.25% 或 117/400。

5.通过将百分比相加计算回答的累积百分比:29.25% + 20.25% = 49.50%。

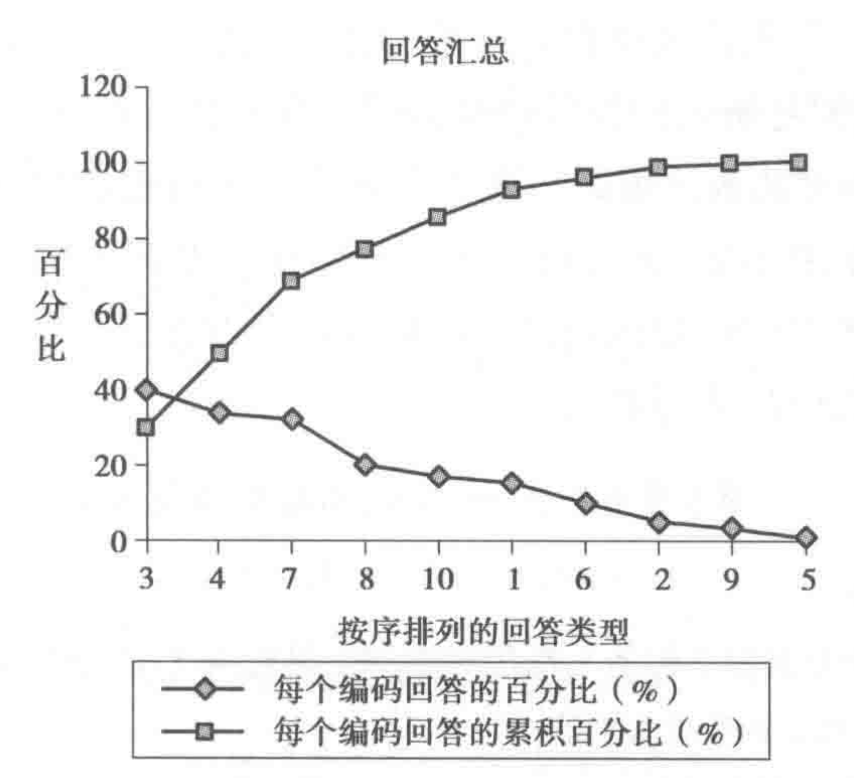
下表是对某个假想数据汇总上述步骤的结果。该表随后通过标准软件转化为线图。线图显示在表后。

■ 示例　　LB/LL 的例子:回答汇总

回答分类(按序排列编码)	每个编码的参加者百分比(100 名参加者)(%)	每个编码的回答数量(100 个参加者)	每个编码的回答所占的百分比(%)	每个编码的回答的累积百分比(%)
3	40	117	29.25	29.25
4	34	81	20.25	49.50
7	32	78	19.50	69.00
8	20	35	8.75	77.25
10	17	30	7.50	85.25
1	15	29	7.25	92.50
6	10	14	3.50	96.00
2	5	10	2.50	98.50
9	3	5	1.25	99.75
5	1	1	0.25	100.00

回答分类(按序排列编码)	3	4	7	8	10	1	6	2	9	5
每个编码回答的百分比(%)	40	34	32	20	17	15	10	5	3	1
每个编码回答的累积百分比(%)	29.25	49.5	69	77.25	85.25	92.5	96	98.5	99.75	100

■



正如你所见,将回答类型按人数多少排序后沿 X 轴分布。Y 轴代表百分比。

对于每一个回答类型,你应该在 X 轴上查看两个点:参加者的百分比和回答的累积百分比。首先,回答的累积百分比是连在一起的方块(▪)。其次,一些代表参加者百分比的点是连在一起的钻石形(◆)。该图显示编码为 3、4 和 7 的回答的频度差不多。这几类是需要强调的,因为最高比例的参与者属于这几类,它们加在一起占所有回答的 69%。

第 8、10 和 1 构成了直观的第二梯队,占所有回答的 23.5%。加起来看,编码为 3、4、7、8、10 的回答占了总数的 92.5%。

撰写封闭问题的原则

多重选择的封闭问题包含一个主干,代表一个问题(通常以一个陈述句、问题、简史或情境的形式出现),接下来是几个可选项或解决方案。以下是构造此类问题的原则。

1.**每个问题对被调查者而言都应该是有意义的**。在一项关于政治观点的调查中,问题应该是关于政治程序、党派、候选人等方面的。如果引入了没有明显意图的问题,比如关于年龄或性别的问题,你可

能需要解释一下为什么提出这样的问题。例如:“我们将提几个个人问题,以便我们考察人们的背景与他们的观点之间的关系……”

2.**使用标准的语法规则**。因为你希望获得每道题的准确回答,就必须使用传统的语法、拼写和句式。避免使用专有名词(除非你希望测试人们关于它们的知识)和缩写,确保题目不过于冗长,否则你实际上是在测试人们的读写能力。

■ 示例 **题目撰写技巧——长度、清晰性、缩写和术语**

长度

差的:在测量人们的下列方面时,缺乏具有高度稳定性和建构效度的心理量表是最令调查者头疼的:

1.经济特征

2.感觉

3.知识

4.健康

较好的:在测量人们的下列方面时,缺少可靠和有效的方法给调查者带来最大问题的是:

1.经济特征

2.感觉

3.知识

4.健康

清晰性

差的:肚子疼时你通常采取何种治疗方法?

较好的:肚子疼时你通常服用何种牌子的药品?

缩写

差的:哪个政党应该对扩大国内生产总值(GDP)负责?

1.共和党

2.民主党

较好的:那个政党应该对 GDP 的减少负责?

1.共和党

2.民主党

术语

差的:依你看来,何种二元关系(dyad)对幼儿时期信任感的影响最大?

1.父亲和母亲

2.父亲及其兄弟姊妹(sibling)

3.母亲及其兄弟姊妹(sibling)

较好的:依你看来,何种家庭组合对幼儿时期信任感的影响最大?

1.父亲和母亲

2.父亲及其兄弟或姊妹(brother or sister)

3.母亲及其兄弟或姊妹(brother or sister) ■

3.**使问题具体化**。问题应该接近于被调查者的个人经验。

■ 示例 **题目撰写技巧——具体的问题**

较不具体:你认为其他人会喜欢这本书吗?

较为具体:你向其他人推荐过这本书吗? ■

与提问被访者是否会向其他人推荐这本书相比,提问他们是否认为别人会喜欢这本书显得更为抽象。谁是“其他人”? 你确信被访者了解其他人对这本书的感受吗? 你的提问与被访者的直接经验越远,你遇到记忆方面问题的可能性越大。

考虑下面的例子:

■ 示例 **题目撰写技巧——问题的针对性**

调查者调查了一座小城的5家公司以了解其对管理岗位雇佣女性的态度。一个调查问题是:“你认为女性获得管理岗位的机会与男性一样多吗?”调查者的一个朋友指出一个更好的提问方式是:“在(填写公司名),女性获取管理岗位的机会与男性一样多吗?” ■

要小心不要在具体方面犯错误。如果你要问人们在过去的一周里他们每天看电视的时间,那么就要确保过去的一周里没有发生什么不寻常的事情,这样数据才能代表从电视观看角度而言的“真实的”一周。电视特别节目,诸如奥运会和坠机、地震、洪水和火灾这样的天灾,是可能影响收视习惯的因素之一。

4.**避免有偏见的词句**。某些姓名、地点和观点可能触发情感上的

挑战。当调查中包含这些内容时,就会影响到人们的回答。诸如总统、流产、恐怖分子和酒鬼这些字眼。

假设你调查刚完成一个节食计划的人们。你使用了下面哪些词:瘦或苗条、肥胖、笨重或多脂?

还记得这些吗?我是坚定的,你是固执的,他是个猪头傻瓜!

看一下这些问题:

你会投票给罗杰·菲尔德斯吗?

你会投票给罗杰·菲尔德斯博士吗?

你会投票给罗杰·菲尔德斯——一个自由主义者吗?

虽然罗杰·菲尔德斯看上去是对候选人最中性化的描述,但也可以被认为是最少诱导性的。而博士或者自由主义者的介绍则会给回答带来偏差。

5.**检查你自己的偏见**。偏见的另外一个来源是调查者没有意识到自己关于某一议题的立场。看看下面的例子。

■ 示例　　**题目撰写技巧——隐蔽的偏见**

差的:你认为自由主义者和保守主义者是否不久后会达成更高程度的相互理解?

差的:你认为自由主义者和保守主义者当前低的相互理解水平是否会持续下去?

■

当你怀疑自己的提问强烈偏向某一边时,请他人来评阅是会有帮助的。请教评阅人你的措辞对某个持相反意见的人而言是否是无偏的和可接受的。对于一个了解人们对自由主义者和保守主义者关系的看法的调查,可以以下面的方式提问。

■ 示例　　**问题撰写技巧——隐蔽的偏见**

较好的:根据你的看法,在接下来的4年中,自由主义者和保守主义者的关系可能会如何变化?

大幅改进

有所改进

有所恶化

大幅恶化

不可预测

■

6.**提问个人信息时注意委婉**。偏差的另外一个来源是可能威胁到被访者的那些问题。诸如"你每年挣多少钱?""你是单身还是离异?"以及"你感觉你的老师、顾问、医生如何"一类的问题是私人性的,可能会触犯某些人而不给出真实的回答。当调查必须获取个人信息时,提问时要以尽可能小的情感挑战的方式提供回答的分类。

■ 示例 **问题撰写技巧——非常个人的问题**

差的:你去年的年收入是多少?

__________美元

较好的:你去年的年收入最符合下面哪一类?

低于 10 000 美元

10 001 到 20 000 美元

20 001 到 40 000 美元

40 001 到 75 000 美元

高于 75 001 美元

■

有敏感问题时更倾向于对回答分类,因为这样可以模糊被访者的身份,使问题不那么个人化。

7.**每个问题只包含一个意思**。以避免被访者的真实回答不能被准确理解。

■ 示例 **问题撰写技巧——每个问题一个意思**

差的:美国是否应该削减其军事或国内开支?

是

否

不知道

较好:美国是否应该实质性地削减其军事开支?

是

否

不知道

美国是否应该把更多的钱用于国内项目上?

是

否

不知道

如果美国削减其军事开支,是否应该把这笔钱用于国内项目?

是

否

不知道

■

封闭问题的回答

是和否

调查中封闭问题的回答有几种形式。

■ 示例 **是和否的回答**

你大学毕业了吗?

是

否

你的车是四轮驱动的吗?

是

否

■

是和否的回答易于使用和计分。但是即使轻微的误解也可能导致回答与被调查者的实际意思完全相反。另外,在某些情况下,绝对肯定或绝对否定的提问可能导致被调查者拒绝回答或选择“不知道”。

评级量表

分类的抑或连续的? 什么是顺序的?

评级量表要求被调查者将被评价的项目置于序列分类中的某一

类或者连续谱中的某个点上。序列分类或者点被赋予一个数值。有三种类型的评级或测量量表。

1.**分类的**。有时也被称为名目回答量表(nominal response scales)。这要求人们确认或命名他们所属的群体:性别、宗教信仰、学校或最后毕业的大学。

■ 示例

分类评级量表

新生儿的性别是什么?

	圈选一项
男	1
女	2

■

2.**顺序的**。这类量表要求被调查者根据重要性次序做出回答。例如一个人的经济地位(高、中或低)就是提供了一种顺序测量。

■ 示例

顺序评级量表

你的最高学历是什么?

	圈选一项
小学	1
中学肄业	2
中学毕业	3
大学肄业	4
大学毕业	5
研究生	6

■

3.**数字的量表**。数字量表有两种形式:离散的和连续的。离散量表会给出一个确切的数字,而连续量表则会给出一个落在连续谱中的数字。

这是离散量表的一个例子:

在过去的12个月中如果你某天喝了酒,通常会喝多少酒(啤酒、红酒、白酒)?

5或更多	□
4	□
3	□
2	□
1	□

回答是离散的,因为你得到的是一个实际的数字。与之不同的是得到一个连续谱中数字的数字量表。

你的血压多高? ________/________mm/Hg

你个子多高(厘米)? ________cm

连续数据可以取任何数值,并且能够根据测量系统的精度被有意义地划分为越来越小的间隔。

你还可以通过图形获得连续数据。图形量表是一个回答的连续谱可视化的评级量表。例如一个关于市议会在解决某个问题上的效率的民意调查,图形量表类似下面的例子。

■ 示例　　**评估市议会效率的图形评级量表**

指导语:你认为市议会在解决以下三个问题上的效率如何?请在线上的相应位置上画一个"×"。

	非常有效		中等	非常无效	
清洁环境	1	3	5	7	9
新的公共交通	1	3	5	7	9
雇用新教师	1	3	5	7	9

■

对于像这样的图形量表,你没有必要命名标度上的所有点。在本例中,只命名了三个点:非常有效、中等和非常无效。

图形量表相对易于使用。注意把描述与相应的点尽可能放得近一点。如果你在同一个问卷的不同部分两次使用了相同的量表,那么

要确保点之间的距离保持一致。

■ 示例

形式不佳的图形量表

非常有效		非常无效		
1	3	5	7	9

■

图形量表的一个主要不足是它们有时难以解释。看看下面三个被调查者对同一个图形量表的标记:

■ 示例

解释图形量表

	同意	介于同意和不同意之间	不同意
被调查者 A		×	
	1	2	3
被调查者 B	×		
	1	2	3
被调查者 C		×	
	1	2	3

■

被调查者 A 明确选择了"介于同意和不同意之间"的评级,但被调查者 B 和 C 呢? 调查者 B 的选择处于"同意"和"介于同意和不同意之间"之间。被调查者 C 的选择处于"不同意"和"介于同意和不同意之间"之间。怎样对这些评级做出解释? 你可以把被调查者 B 和 C 归入最接近的评级点——在本例中就是 2,或者你可以假想一个真实的连续谱,赋予被调查者 B 以 1.75 的评分和被调查者 C 以 2.20 的评分。因为所有的图形量表都面临类似的解释困难,调查者总是需要确定一种策略以使被调查者的评级的意义明朗化。

分类、顺序和连续数据的区分是重要的,因为它们决定了你所能采用的统计处理方法。例如平均数。如果你想得到连续数据的平均分数,你会用到算数平均值或均值。而对于顺序数据,你可能考虑计算中位值或将所有回答划分成高低两半的数值(更多关于统计方法和

调查数据的内容参见第 6 章)。

顺序量表

当评级者使用顺序量表时,他们在具有某种顺序的有限分类中做出选择。

■ 示例　　　　顺序量表

- 经常
  有时
  几乎从不
- 非常喜欢
  喜欢
  既不喜欢也不厌恶
  厌恶
  非常厌恶
- 非常同意
  同意
  既不同意也不反对
  反对
  强烈反对
- 完全同意
  基本同意
  既不同意也不反对
  基本不同意
  完全不同意

■

顺序量表易于使用和解释。以多少类为宜?有些人多至 9 类,另外一些人少至 3 类(高、中、低)。双数的选择(比如说 4 类)会迫使被访者不能持中间立场(不置可否)。但是分类的数量一定要根据调查的需要和被调查者的技能来定。如果需要准确的信息,并且被调查者有意愿和能力提供,你也有资源收集这些信息,那么就可以使用多一些的分类(7 到 9 类),否则的话分类数就少一些。

考虑下面两种情形:

■ 示例　　　　选择分类数量

1.开展一个4分钟的电话调查,了解有全职母亲的家庭在餐馆就餐的频度。提出的问题是:“在一个一般的月份里,你的家庭多久在餐馆吃一次饭?”回答选项包括每周两次及以上、每周一次或一周少于一次。

2.要求医生对开展冠状动脉搭桥手术和胆囊切除之类的外科手术的一系列理由的适当性做出评级。一个适当的理由被界定为对病人而言收益大于风险。采用了1到9的刻度,其中1=完全不恰当,而9=完全恰当。

■

在情形1中,4分钟的访谈只能给出简短的回答。在情形2中,则要求医生根据他们的专业知识做出相当精确的评价。

一种经常使用的顺序量表类型是李克特量表,要求被调查者说出他们对一个表述同意或不同意的程度。刻度一端的数字代表最不同意或强烈反对,刻度另一端的数字则代表最同意或强烈同意。李克特量表通常分5级:

强烈反对

反对

既不同意也不反对

同意

强烈同意

有时候也使用4级量表。这是一种强迫选择方法,因为没有提供“既不同意也不反对”或者常见的“中立”的中间选项。根据调查需求的不同,选项可以从“强烈反对”到“强烈同意”,也可以反过来排列。如果在量表的一端有其他词语进一步明确数字的意义,那么这被称为李克特类型的问题。

你认为标准化测验分数对五年级学生教育的重要性有多高?				
不很重要				极其重要
1	2	3	4	5

李克特问题的回答有时被当作顺序数据加以分析,特别是当调查者不能假定被调查者意识到相邻等级的差距是等距之时。如果被当作顺序数据,李克特回答可以制成条形图,以中位值或众数(而不是平均值)汇总,以变动范围(而不是标准差)反映变差,采用非参数检验(如卡方检验或 Mann-Whitney 检验)进行分析。如果调查者将其视为连续的,则意味着标准差是恰当的统计量。当调查者希望比较平均分数时可以将数据视为连续的。来自李克特量表的数据有时也通过将所有的“同意”和“不同意”的回答归并成“接受”和“拒绝”两类来减少分类层级。做了这样的转换后,卡方和 McNemar 检验是常用的统计方法。

如果你希望使被调查者不能那么随便地选择中间类别(介于同意和不同意之间的不置可否)或敷衍回答(不肯定,不知道),那么强迫选择性问题经常是有用的。但是强迫被调查者做出选择可能惹恼他们,导致不能发现他们真实的想法。

有些顺序量表依赖相对的判断和评级,如下例所示。

■ 示例

等级次序量表

请根据他们的写作能力对下面的五个人进行排序。最高等级的赋予数字 1,最低等级的赋予数字 5。

______________费伊·格罗斯

______________贝蒂·巴斯

______________爱德华·罗姆尼

______________亚历山大·拉尔曼

______________马文·杰克逊

■

排序涉及因素之间的比较。顺序量表的另外一种形式是比较量表。该量表要求被调查者将一个特定对象与一般水平进行比较。

■ 示例

比较性评级量表

请比较泰皇餐厅与洛杉矶的其他餐厅。选择一项:

☐它比大多数都好

☐它和大多数差不多

☐它不如大多数那么好

■

为了保证等级顺序和比较性评级量表提供准确的信息,你必须确

信被调查者能够做出比较。他们是否有判断写作能力的经验?如果没有,他们如何才能排序?他们对洛杉矶的餐厅是否足够熟悉?如果不是他们如何才能做出比较?

选单

选单为被调查者提供了一系列答案。他们可以根据提示选择其中一项或多项。

■ 示例 **被调查者必须从多个回答中选择一项的选单**

你认为下列药品中最适合治疗头疼的是哪种?
(只选一项)

	圈选一项
我头疼不吃药	1
阿司匹林	2
退热净(羟苯基乙酰胺)	3
艾德维尔(Advil)	4
伊克赛锭(Excedrin)	5
其他,请指明________	6

■

■ 示例 **被调查者就列表中的每一项做出是、否或不知道回答的选单**

在每一行的一个框内打钩

在过去的12月中,是否有医生或其他医务工作者告诉你患有下列病症:

	否	是	不知道
a.肝炎	□	□	□
b.胃炎	□	□	□
c.胃或小肠溃疡	□	□	□
d.胰腺炎	□	□	□
e.抑郁、焦虑或其他精神、心理方面的问题	□	□	□

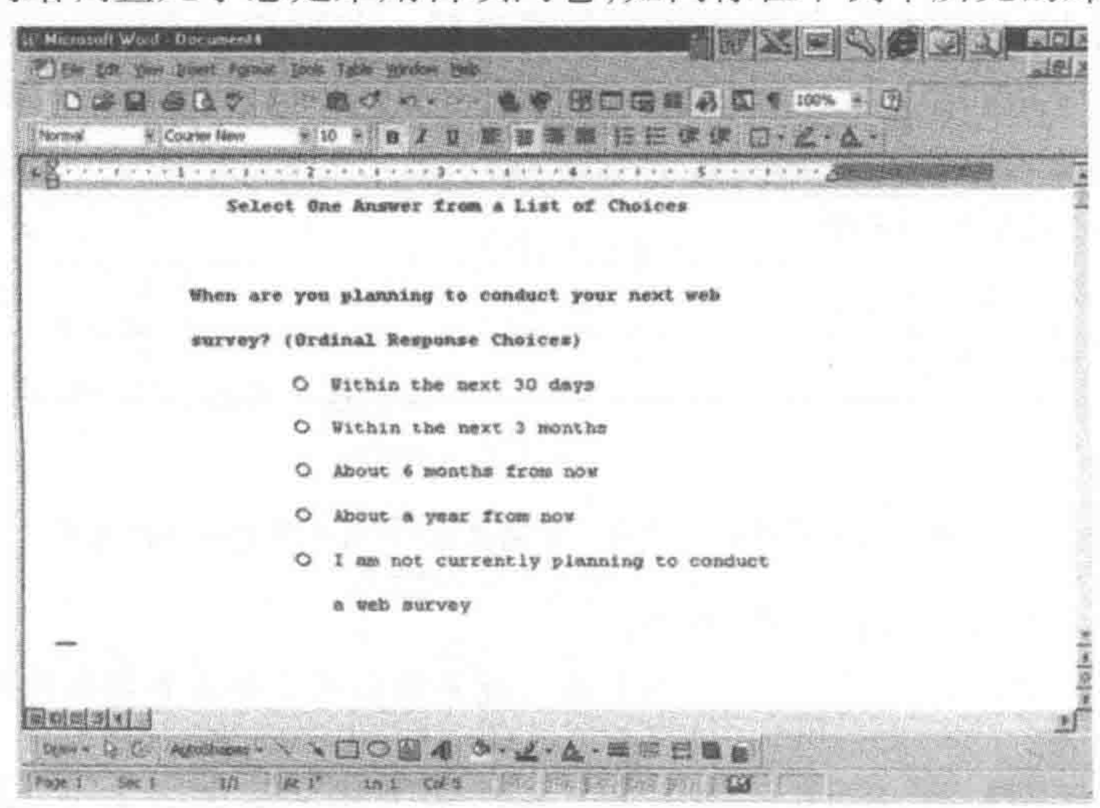
■

选单能够提醒被调查者某些他们可能忘记了的事情。如果你只

是简单地要人们列出他们吃过的药,可能有些人会不知道。选单还有助于解决写不出难词生词带来的问题。选单的问题之一是被调查者可能觉得某个选项很熟悉而选择了错误的答案。他们服用的是盘尼西林(penicillin)还是氨苄青霉素(ampicillin)(两个单词的拼写相似)?是这个月还是上个月?可以做出多重回答的选单的形式和解释会比较麻烦。假设在第一个例子中,某人选择了阿司匹林,但却没有说明是否服用了其他药物。你如何能肯定他就没有服用其他药物?也许他服用了但却懒得做出完整回答。

网络调查

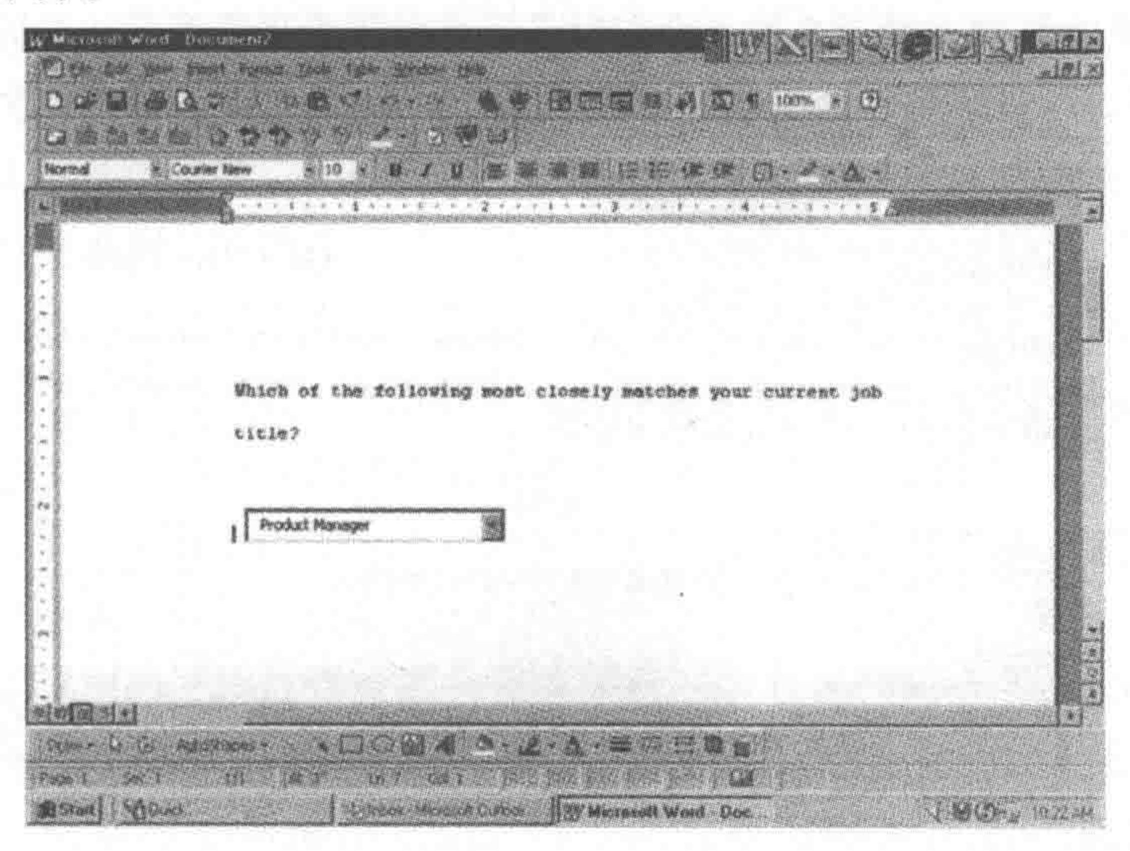
网络调查几乎总是采用自填问卷,如同你在下例中所见的那样:



网络调查创建和问题选择的原则与其他自填问卷几乎完全相同。调查必须被设计得可以从最有可能做出最准确回答的人那里获得所需要的信息。问题必须语法正确,选项的范围必须合理。

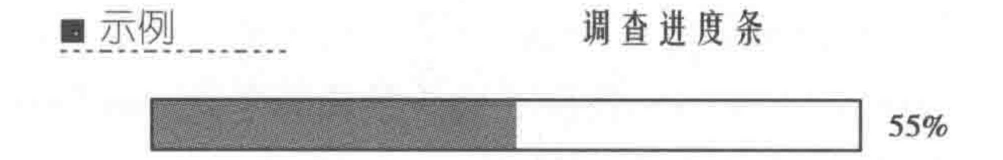
下拉列表是网络调查的一种独特的问题形式。当回答的列表可能很长时,经常用它来节省空间。例如,如果你希望知道被访者生活在哪个城市中,通过提供城市的下拉列表,就会比完整显示它们节省很多屏幕空间。

下拉列表经常用于节省被调查者的时间以及确保回答的标准化。比如下图这个问题,询问被调查者在一个虚构的有数百名员工的公司中的职务。



网络调查经常包含一个进度条。进度条告诉被调查者他们已完成的调查所占的比重。这是一个例子。

■ 示例



■

进度条告诉被调查者完成调查实际所需的时间。如果被调查者用 10 分钟完成了 20% 的调查,那么他们将需要 50 分钟才能完成 100%。

有些在线调查设计允许被调查者随时登入登出。如果被调查者需要 50 分钟完成调查,但是此时他们只有 30 分钟时间,登入登出功能能够鼓励他们在其他时间返回调查。对于网络连接可能中断的调查参与者而言,让他们能够随时登入登出是个好主意。对于那些登出后忘记了他们还未完成调查的被调查者,得准备好向他们发送邮件提醒。

网络调查问题可以根据对一个问题的回答定制下一个问题的回

答,如下例所示。

■ 示例

你使用哪家网络接入商?

○ AOL

⊙ MSN

○ Earthink

○ Quest

请对××××的下列方面做出评价:

	很好	好	中等	差
联通性	○	○	○	○
速度	○	○	○	○
客户服务	○	○	○	○
总体表现	○	○	○	○

请对 MSN 的下列方面做出评价:

	很好	好	中等	差
联通性	○	○	○	○
速度	○	○	○	○
客户服务	○	○	○	○
总体表现	○	○	○	○

来源:http://www.questionpro.com/help/162.html ■

网络调查应该解释如何退回修改回答,然后再继续前进。考虑使用下面的指导语:

请在开始调查前阅读下列指导语。

前进:

点击页面底部的 NEXT 按钮以保存你的回答并进入下一页。

退回:

使用页面底部的 PREVIOUS 按钮回看上一页的回答。你可以修改你以前输入的回答。不必使用浏览器的 BACK 按钮返回上一页。使用浏览器的 BACK 按钮会导致你退出调查。如果出现这种情况,你需要等待 10 分钟才能使用电子邮件中提供的链接再次返回调查。

平实且简单的调查问题和回答

调查问题和回答选项的文字最好适合6到8年级的阅读水平(基于美国的学制体系,相当于中国的初中)。为什么?研究显示一般的美国人在这个水平上阅读最舒适。事实上,有证据表明大多数人更喜欢使用简单词句的调查。

有许多办法可以确定一份文本的大致阅读水平,从网上可以找到这些方法的相关信息。大多数可读性测验提供一个年级水平分数,该分数基于每个词包含的平均音节数量和每个句子的词数。一个单词中的音节数量越多以及一个句子中的单词越多,阅读和理解该文本的难度也越大。

最常用的可读性测验之一是弗莱什—金凯德(Flesch-Kincaid)测验。每种方法都各有优缺点,专家们对哪种更好也没有一致意见。你可以上网学习如何使用它们。可以从这些关键词开始:"阅读能力统计"。

弗莱什—金凯德方法被认为不如弗赖伊(Fry)法准确,但后者更耗时。你可以通过使用 Microsoft Word 提供的可读性工具很容易地计算出弗莱什易读性分数。以下是方法。

1.进入"审阅"标签。

2.进入"拼写和语法"标签并点击"选项"。

3.在"语法"标题下选择"显示可读性统计量"。

每次做拼写和语法检查时,都会同时被给出几个可读性统计量,包括弗莱什—金凯德年级水平分数。

量 表

加成量表(Additive Scales)

大多数调查都是对每个问题分别计数。在一个调查人们对生活

在拖车公园的态度的调查中,你可能会提出 12 个问题,每个问题都是关于态度的,分别计分。假设你收集下面这样的信息:

- 生活在拖车公园中的时间
- 之前是否在拖车公园生活过
- 对拖车食宿条件的满意度
- 对公园食宿条件的满意度
- 对生活方式质量的满意度
- 拖车的舒适性
- 拖车的车龄
- 汽车的车龄
- 汽车的类型
- 被调查者的年龄
- 被调查者的性别
- 年收入

通过这些信息,你可以分别报告每项事实,或者考察以下关系:

被调查者年龄与生活方式质量满意度之间的关系。

性别和在拖车公园生活时长之间的关系。

然后,其他调查可能与此不同,每个问题不是分别计算,而是结合在一起形成一个分数。

考虑下面的例子:

■ 示例 **有加成量表的调查**

某大学医疗中心的医生观察到他们的很多重病人在社会生活中表现良好。尽管有残疾,但他们有可能可以一起去看电影、购物等。同种疾病的其他病人待在家里,与自己的朋友和家庭隔离。医生们假设这两组病人之间在心理功能和个人生活便利资源的获得能力方面存在差别。作为检验他们假设的一部分,他们决定对两个群体做赫伯特医学院和加利福尼亚大学联合开发的功能状况调查(Functional Status Inventory)。经过 5 年的研究和验证后,两所大学的研究者准备了一个供慢性病人使用的功能调查。高分意味着良好的功能状态,低分意味着不佳的功能状态。 ■

制作加成量表所使用的方法需要较高水平的调查研究技能,因为

你必须确切证明在全部的每一个方面高分者确实不同于低分者。

制定加成量表

调查者至少在两种意义上使用"scale"一词。第一种指的是答案的组织方式:

1.你是否每天吃六份或更多水果或蔬菜?

回答标度如下:

是 1

否 2

2.你对本书中例子的满意程度如何?

回答标度如下:

非常满意	1
比较满意	2
不太满意	3
很不满意	4

对"scale"一词的第二种使用方式指的是其分数有意义的一个问题或一组问题。如果你有10道关于生活质量的调查问题可以计分,1分表示低质量,10分表示高质量,这就构成了一个生活质量量表。

■示例 **外语技能调查**

圈选最能描述你说下列语言能力的类型。

	流　利	比较流利	不流利
法语	2	1	0
德语	2	1	0
意大利语	2	1	0
西班牙语	2	1	0
瑞典语	2	1	0

■

对于每一项——法语、德语等——采用了等级序列量表组织回答。同时,将五项相加,就能得到一个关于语言能力的量表。对所有

语言都流利的被调查者位于量表的一端，对所有语言都不流利的位于另一端。假设你赋予对每种语言流利为 2 分，比较流利 1 分，不流利 0 分。一个对所有语言流利的人将得到最高的 10 分，而对所有语言都不流利的将得到最低的 0 分。在本书中，这种类型的量表被称为加成的（因为对每个问题分别的回答将被加总在一起，或者说结合在一起）。最常用的加成量表是区分的、总和的和累积的。

区分量表

区分量表用来分辨是否同意某一理论或观点的人。例如，为了制定针对某一概念（比如机会平等）的区分量表，意味着需要组合很多陈述（如"有能力的男人和女人应该同工同酬"），并且要求专家就每条陈述是否支持该概念做出评价。然后你计算专家对每个陈述的平均或中位评级。接着你询问被调查者同意还是不同意每条陈述。他们的分数仅取决于被调查者表示同意的那些问题。为了得到被调查者的分数，你查看被调查者选择的每条陈述对应的专家平均分，把这个对应分加起来，然后计算算术平均值。

给区分量表使用者的典型指导语有点类似这样：

请勾选你同意的每一条陈述。

或

请勾选最接近你立场的条目。

区分量表的计分采取下面的方式。

■ 示例　　区分量表的计分

学生 A 接受了健身调查，被要求选出她最同意的两个或三个条目。这是她选出的两个条目以及专家的分数。

专家给出的中位分数	
1.健身是一种日渐兴起的观念	3.2
2.定时运动（如走路或骑行）对每个人都是必要的	4.0

■

学生 A 的分数是 3.6（3.2 和 4 的平均值），这个分数可以被认为

支持健身。(可能的最好分数是 1.0,最差分数是 11.0。)

区分量表有没有缺点?或许它最大的缺点是建构量表所需的工作量很大。另外,你还必须考虑标定量表中分数和解释回答定义的专家的态度和份量,专家可能与使用量表的人非常不同。

总和量表

总和量表根据人们的回答加总进行排序。假设一个自尊问卷有一系列问题,使用相同的评级标度(同意、中立、不同意)。

■ 示例　　**制订一个自尊调查所用的总和量表**

指导语:对于下列每一项陈述,请勾选你是同意还是不同意。

	对每个陈述在一个框中做记号		
陈　述	同　意	既非同意也非不同意	不同意
a.有时我觉得我一无是处	□	□	□
b.总的来说我对自己满意	□	□	□
c.我经常觉得非常孤独	□	□	□
d.我的社会生活十分完整	□	□	□
e.我的朋友赞佩我的诚实	□	□	□

■

对于这份问卷,你如何计算总和量表?首先,确定哪些问题是正向的(在本例中是 b、d 和 e),哪些问题是反向的(a 和 c)。接下来,给每类回答赋予一个数字权重。你可以这么做:

正向的=+1 分

居中或中立=0 分

负向的=-1 分

某人的分数是他或她对五个问题回答的算数和。某被调查者 X 的回答见下例所示。

■ 示例

总和量表的计分

陈　述	被调查者 X 的回答			问题是正向的(+)还是负向的(-)	问题得分
	不同意	中立	同意		
有时我觉得我一无是处	✓			-	+1
总的来说我对自己满意		✓		+	0
我经常觉得非常孤独			✓	-	-1
我的社会生活十分完整			✓	+	+1
我的朋友赞佩我的诚实		✓		+	0

■

X 不同意问题 a,该题基本是负向的,因而得到+1 的分数。对于题 b,X 保持中立,因而得分为 0。题 c 是同意,但 c 基本上是负向的,因而得到-1 分。题 d 是同意,得到+1 分。对题 e 的回答是中立,得分为 0。X 的总和量表得分为+1,该量表最大可能得分为+5。(如果 X 的回答是:题 a=不同意;题 b=同意;题 c=不同意;题 d 和 e=同意,那么就能得到最高的+5 分。)

小　结

调查的内容取决于对术语的清晰定义,因为人类的态度和感觉在不同理论中描述是不同的。

确保被调查者能够回答问题。他们是否足够熟悉这个主题?

不要去问你用不到的信息,因为这是对被调查者和你自己时间的浪费。

分类评级量表是请人们确认他们的群体归属,包括性别、宗教信仰、学校或最后上的大学。

调查问题可以是封闭的或开放的。有若干选项的封闭问题比开放问题更容易计分。

开放问题给被调查者用自己的话表达立场的机会。不幸的是,这

些话解释起来很难。

撰写问题时,使用标准的语法和句式。尽量使问题具体化并接近被调查者的经验。字词、名字和观点可能不知不觉地导致你的结果发生偏差,对此要保持警觉。不要过于个人化。每个问题只包含一个意思。考虑采用某种可读性测试。

数字型量表产生离散的和连续的两类数字。前者产生一个准确的数字,而后者产生一个落在连续谱中的数字。

顺序量表要求被调查者将回答置于重要性的序列中。

一种经常使用的顺序量表是李克特量表,它要求被调查者说出他们同意或者不同意某一表述的程度。

网络调查提供特殊的问题与回答形式,如下拉菜单。很重要的一点是确保被调查者知道如何在网络调查中行进,特别是如何前进和后退(以便修改回答或查阅)。

思考一下

1.阅读以下两组问题。说出它们是否使用了分类、顺序或连续数据量表。

A.有关网上搜寻健康信息的问题

当你搜寻健康信息时一般会上大约多少个网站?	
□	1
□	2 或 3
□	4 或 5
□	6~10
□	11~20
□	超过 20
□	不确定
你上这些网站是通过网络搜索还是通过某些其他方式找到它们的?	
□	通过网络搜索

续表

□	通过其他方式找到
□	不确定
你之前是否听说过这些网站?(如果是,提出以下问题:你是在哪里听说的。新闻文章、广告、医生或医生办公室、健康保险公司、朋友或亲属还是其他地方?)	
□	新闻文章
□	广告
□	医生或医生办公室
□	健康保险公司
□	朋友或亲属
□	其他地方

B.关于饮酒的问题

在过去的12个月中,你饮用含酒精的饮料的频率如何?	
涂黑一个圈	
○	每天或几乎每天
○	每周4或5次
○	每周2或3次
○	每月2到4次
○	每月一次或更少
○	从不——请转到第×××页
在过去的12个月中,大约有多少天,你在饮酒3杯以上后驾驶汽车、卡车或其他机动车辆?	
涂黑一个圈	
○	20天或更多天
○	10~19天
○	6~9天
○	3~5天

续表

○	1~2 天
○	从不
○	我过去 12 个月中未驾驶过车辆
你如何描述你现在的健康状况?	
○	非常好
○	较好
○	好
○	一般
○	不佳
你是男性还是女性?	
○	男性
○	女性

2.准备一个少于 10 题的自填问卷,了解学生们对本章内容的满意度。分类、顺序和数值数据,每类至少包含一个问题。对每个问题的选择及含义进行说明。

3.解释这些术语:开放问题,封闭问题,强迫选择型问题,等级次序量表,比较评级量表,李克特量表,下拉菜单,总和量表和区分量表。

3

调查研究的一些实践要点



一个调查应该持续多长时间？调查的长度取决于你需要了解什么、你什么时候去了解、被调查者愿意用来回答问题的时间和你的资源。

调查中的第一个问题应该明显与调查目的相关；客观问题应该出现在主观问题之前；从被调查者最熟悉的开始逐渐过渡到最不熟悉的，并遵守自然的时间流程；保持问题的独立以避免偏倚；将相对容易的问题放在最后（特别是在长调查中）而将"敏感"问题放在中间；避免许多相似的问题；按逻辑顺序排列问题。

为了提高自填问题的回答率，需要事先计划并监控结果。网上调查系统能够让你实时做到这一点。考虑在调查开始前以及调查期间向被调查者发送邮件。可以是普通邮件或电子邮件。建议向被调查者发送调查结果；使问卷简短，并考虑提供奖金。访谈需要好好准备并做计划。访谈员需要系统的大量培训。建立一个体系监督访谈全程的质量。

调查应该做小规模测试，以检验调查能否很容易地被接受和执行以及是否符合计划。你的主要目标是可靠和有效的调查数据。可靠性指的是获取的信息的一致性，有效性指

的是信息的准确性。

提高调查的可靠性和有效性的方式之一是:使其以一个已被证明有效的调查为基础,但要谨慎。寻找调查可靠性和有效性的证据。

小规模测试有助于你设计一个可靠的调查。设想调查实际发生的环境并制订计划加以应对。选择与最终参与调查相似的被调查者。对于可靠性,关注问题的清晰性和调查的一般形式。

在会遇到多种语言和多种文化的环境中,调查者要了解如何请语言流利的人帮助翻译调查,并检测问题和回答是否可以准确表达。

小规模测试之所以能够提高可靠性和有效性,是因为它能帮助你检查是否包含了所有的主题以及充分的问题选项——如果人们确实不同,你的调查应该考虑到这些差别。

大多数开展调查(作为独立研究或较大研究活动的一部分)的公共或私立机构都有自己关于保护被调查者隐私和秘密的规则。准备好向被调查者解释他们完成调查所面临的风险和收益。

长度方面的考虑

调查的长度取决于你需要了解什么,以及最少需要多少问题才能让相应的回答可信。另外一个考虑因素是被调查者有多少可用的时间,以及他们对调查的注意力如何?例如,较小的孩子只能集中注意力几分钟。你还需要考虑你的资源。一个 10 分钟的电话或面对面访谈的花费要少于 20 分钟的同类访谈。网络调查轻轻一点就能发送给成千上万的人,但设计网络调查所需的专业技术知识是昂贵的,并且如果你和一家商业软件供应商签约,可能需要付出版权费并签署特殊的隐私协议。

以下示例两种情境,说明调查开展的环境如何影响问卷长度。

■ 示例　　**调查的环境如何影响问卷长度**

情境 1:本地图书馆希望确定它是否仍能满足变迁中的社区需求。近年来,它的用户更多的是 65 岁以上的老人,很大一部分还是以英语作为第二

语言的。图书馆的主要关切包括:他们的展览、报纸、杂志和其他定期出版物、关于新书和作者的项目、与健康保健问题相关的特别收藏,是否充分和有用。一位双语志愿者将每周花费两个上午和一个下午,一共 8 周,开展一项面向图书馆用户的 45 分钟时长的面对面访谈。为此设计了一个 50 题的调查表。

情境 2:社区图书馆还想确定他们的服务是否适合日渐增加的老年和非英语人口。该市认为并非只有社区图书馆面临需求的变化,因而同意开展一项关于图书馆用户的调查。为了尽量减少图书馆工作人员和用户花费的时间,由中心图书馆办公室准备了一个 10 分钟 6 题的自填问卷。4 题询问的是图书馆的书刊收藏、计算机接入和特别收藏的充分性。2 题询问被调查者的教育背景和收入。为了方便调查,问卷放置在服务台,被调查者在馆内完成,交给本地图书管理员后再由他们寄到中心办公室做分析。■

在第一种情境中,图书馆付出了相当多的时间去开展访谈以了解社区需求。在第二种情境中,调查则相对简短。

调查要井然有序

所有调查都应该以指导语开始,第一组问题应该是与指导语中描述的主题相关的。看看下面这个调查访谈的指导语和第一个问题。

■ 示例　　**一个调查的指导语及其第一个问题**

你好,我从加利福尼亚大学致电给您。我们正在调查住在学生宿舍的人,了解他们的居住满意度。你的名字是从住宿登记中随机选出的,住宿登记是学生自愿提供了电话号码的名录。我们的问卷不会超过 4 分钟。您可随时打断我。我可以开始提问了吗?

[如果是,继续。如果否,说"谢谢你,抱歉打搅",然后挂断电话]

[这里继续]

第一个问题,请问你对宿舍的整体满意度,你认为[读出选择]

非常满意

比较满意

不太满意

非常不满意

[不读]无意见或不知道/错误回答 ■

访谈者一开始就说将问一些关于学生宿舍满意度的问题,因此第一个问题就是对满意度做出评价。当第一个问题询问客观事实时,人们的反应常常是最好的。一旦他们适应了调查并逐渐了解调查目的,可能更愿意回答相对主观的问题。举例来说,假设你想了解一个暑期城市清洁项目是否成功。你可以首先问被调查者他们最早从哪里听说项目以及参与项目多少时间(两个关于事实的问题),然后再问他们是否喜欢他们的工作。

问题应该从最熟悉的到最不熟悉的。在一个关于健康服务需求的调查中,可以首先问被调查者自己对于服务的需求,然后再问社区的、国家的,等等。

回忆性的问题应该按照自然时序组织。不要泛泛地问"你最早从什么时候开始喜欢慢跑"或"为什么你选择慢跑而不是其他健身项目"这类的问题,而是提示被调查者:"你高中时是否对慢跑有一定兴趣?大学时呢?"

有时候对一个问题的回答会影响另一个问题的内容。当出现这种情况时,就会严重削弱问卷的价值。看下面:

■示例 **排序调查问题**

应该先提问哪个问题?

A.你的导师给了你多少帮助?

B.在你的学业方面你最希望的改进是什么?

答案:应该在问题A之前提问问题B。如果不这样的话,就会不恰当地强调师生关系,仅仅是因为这一点刚刚被提到过。 ■

下面是另一个例子。应该先提出哪个问题?

■示例 **排序调查问题**

A.你对总统的经济政策满意程度如何?

B.总统的领导力如何?

答案:问题B应该先于问题A,因为对总统的经济政策(可能除此之外没有其他不满)不满意的人可能会对总统领导力做出比不这样提问时相对

较低的评价。 ■

把相对容易回答的问题置于调查的末尾。当问卷长或内容较为困难时,被调查者可能会变得疲劳,回答最后的问题时会比较草率甚至干脆不答。你可以把人口学问题(年龄、收入、性别和其他背景特征)放在末尾,因为它们能够被快速回答。

避免许多相像的问题。20 个问题都是询问被调查者同意还是不同意,会导致疲乏和厌倦,被调查者可能会放弃答题。为了尽量减少兴趣的损失,将问题分组,并提供描述形式或主题的过渡。

越敏感的问题越要往后放。诸如打扮习惯、宗教观点和在诸如流产和控枪一类争议问题上的立场,应该适当靠后,这样既能引起被调查者注意又不会因为过于疲劳而不能正确回答。

最后,问题应该符合逻辑顺序。不要从一个主题跳到另一个主题,除非你提供了过渡说明以帮助被调查者理解顺序的意义。

以下是在你的调查中确定问题顺序时所要考虑的要点的清单:

[指导问题顺序的清单]

- 对于任何给定主题,相对客观的问题应先于主观问题。
- 从最熟悉的到最不熟悉的。
- 遵循自然的时间顺序。
- 务必使所有问题相互独立。
- 相对容易回答的问题应该在最后问。
- 避免许多相似的问题。
- 敏感问题不要放在调查开始不久或接近结束时。
- 问题应该按逻辑顺序排列

问卷形式:美观及其他注意事项

问卷的外观非常重要。难以阅读的书面自填问卷会令被调查者感到困惑或烦恼。结果就是数据的损失。没有为记录答案留出足够的空白,这样设计不良的访谈表会降低哪怕最熟练的访谈人的效率。

以下是一些注意事项:

可以:一行只放一个问题,为回答留出足够空间。

不可:把几个问题挤在一起。不要简化问题。

回答形式

一般原则是为做适当记号留出足够空间。以下是几种回答形式。

■ 示例　　　　回答形式

A. ________1.是

　 ________2.否

　 ________3.不知道

B. [1].是

　 2.否

　 3.不知道

C. 　　　　编码

　 是　　　　[1].

　 否　　　　2.

　 不知道　　3.

如果你采用形式 A,注意前面留出足够空间做记号,形式 B 或 C 相对安全一些。 ■

分支问题,或声名狼藉的"跳转"模式

当你关心的问题只适用于部分被调查者时会发生什么?假设你正在做一个关于年轻人参与课外活动的调查。你知道一种主要的活动是运动,另一种是音乐,但你也知道只有部分人同时参与两者。

如果你想问一个你事先已经知道不是与所有被调查者相关的问题,你可以设计一个类似下面例子的表格。

■ 示例　　　　跳转模式或分支问题

3.你是否参加运动?

　 a.否(转至问题 4)

　 b.是

如果是的话,你参与何种运动?

	是	否
足球	1	2
田径运动	1	2
其他,请指明:		

或者

你是否参加运动?

是(完成A部分)

否(转至B部分)

在书面自填问卷中使用分支问题(也常被称为跳转模式)必须格外小心。事实上,有些调查者认为这会令绝大多数人感到困惑,因而完全不应该在自填问卷中使用。记住,如果被调查者不清楚转至调查表的何处,就会遗漏重要的问题。同时也要训练访谈者了解分支部分,不然他们就不能正确开展调查。

网络调查是处理分支(有时也称作“管线”)的有效工具,因为你可以设计软件指导被调查者,并且几乎每个商业供应商都为使用跳转逻辑提供了便利。例如,如果调查逻辑是“如果被调查者对问题的回答是否,那么就转到问题6,”对问题3的回答为否的被调查者就被自动且轻松地(从被调查者的视角看)转到问题6。

执行:谁给谁什么

自填问卷

自填问卷需要大量的准备和监督才能获得满意的回答率。这些问卷被直接交给被调查者,当他们不理解某个问题时通常很少能得到帮助。例如,一张关于教师在职培训的有用性和需求的调查问卷可能是放置在他或她的邮箱里,要求一周后返回。当然,对调查表感到困难的教师可以相互讨论,但并不能保证结果比一个人时更好。

事先准备,通过仔细的编辑和试用,无疑有助于制定清晰且可读

性强的自填问卷。为了进一步保证你能获得你需要的东西,你应该检查返回的问卷。你是否得到了你期望的回答率?是否所有的问题都得到了回答?下面是使用自填问卷的清单。

[使用自填问卷的清单]

- 通过普通邮件或电子邮件事先发信给被调查者,告诉他们你的问卷的目的。提醒被调查者调查即将开始,解释为什么被调查者需要回答问题,并且告诉他们被调查对象是谁。
- 与问卷一同提供一个简短正式的说明。如果你已经通过信件或电子邮件通知了被调查者,那么你可以简化说明。它应该再次描述调查目标和被调查者。
- 提出向被调查者寄送研究发现的摘要。你可以设计网络调查自动显示结果。(如果你对此作出承诺,要安排相应预算。)
- 如果你提出一些被认为是个人化的问题——诸如性别、民族/种族、年龄或收入,要解释为什么这些问题是必要的。
- 力求调查程序的简单。对邮件调查提供已经贴好邮票、写好地址的信封。把纸叠到最小,让被调查者不感觉这是个复杂的力气活。
- 使问卷简短。只问你确实需要的问题,而且不要把它们挤在一起。在书面调查中,给被调查者留下足够的空白以便书写,并且保证每个问题都相互隔开。
- 考虑奖励。这能鼓励人们作答。奖励可以是钱、邮票、笔、食品和免费停车,等等。
- 准备好跟踪或寄提醒信。提醒信要简短且切中要害。不要忘了为寄信和跟踪电话留出预算和时间。如果你使用网络调查软件,确定你的合同中包含跟踪的选项。
- 如果你正在做网络调查。以下是一些特别的注意事项。
  - 在你上网创建调查前准备好所有的调查问题和回答。调查软件会指导你问题的类型(如有一个答案的单选题或

"勾选所有适合的"多选题),但是不要让自己仅仅是为了方便而受到调查软件的影响。

- 如果必要,为完成每个问题提供指导语。
- 确定调查是连续的文档还是分成每屏一个或多个问题。如果你选择每屏一个或多个问题,确保每屏都具有可读性。
- 在网上查看每个问题。如果难于阅读或使用,请重新编辑。如果你把所有问题放在一个文档里,检查文档以确保问题的正确顺序和易于完成。
- 让被调查者随时了解他们的进度。网络调查软件允许你提供进度图示。

访谈

寻找访谈人。访谈人应该尽可能让被调查者容易接纳。他们应该避免怪异的服装、发型等。有时候选择熟悉被调查者的访谈人是个好主意。例如,如果你想了解年轻女孩为什么吸烟,你可以雇佣年轻女性去提问。

同样重要的是访谈人应该说话清楚且善解人意。不寻常的说话方式或口音会带来不必要的麻烦或不希望得到的反应。在调查访谈中,访谈人的说话方式是一个极其重要的需要被考虑的方面。你应该清楚认识到访谈人对调查和被调查者的态度会影响结果。如果访谈人期望不高并释放这种信息,可能会影响到回答率。为了确保你获得尽可能准确的信息,需要对访谈人进行监督。

培训访谈人。好的电话访谈或面对面访谈的关键是培训,以便在相同的时间内让所有访谈人了解对他们的期望,以及让他们以同样的方式提出所有问题。

无论你是培训2个还是20个访谈人,重要的一点都是找到时间把他们聚在一起。会议的好处是可以使用一套标准话语并分享问题。如果受训人需要出差赶来,你需要考虑支付汽油或其他私人或公共交通费用。

一旦进入培训场所,受训人需要有足够的空间就坐和写,或者做你要求他们的任何活动。如果你希望模仿真实情景相互访谈,要保证房间足够大,以使两组或者更多组人说话不会相互干扰。你或许需要几个房间。

如果培训时间超过一个半小时,你需要提供某种形式的茶歇。如果你负担不起,至少留出时间让受训人课间休整。

受训人应该逐步熟悉他们的任务,并且被给予提问的机会。要给他们讲一讲任务的意义,以便他们对可能的困难有所预见并去设法加以解决。确保受训人获取了所有他们工作所需信息的最有效率的方式是准备一个手册。其中,你可以解释他们要做什么,什么时候、什么地点、为什么以及如何做。考虑建立群邮件或在项目网站上设立一个栏目记录和共享调查步骤和政策的改变。

开展访谈。以下是一些在培训会上应该强调的开展访谈的要点:

[访谈培训要点清单]

1.做一个简短的介绍性说明,描述谁进行的访谈("来自阿姆斯特朗纪念医疗中心的玛丽·多伊"),为什么开展访谈("了解你对我们医院项目的满意度"),解释如何选择被调查者("从过去 12 个月出院的人中随机抽取的样本"),以及是否会对回答保密("没有您的书面许可不会使用您的名字")。
2.尝试让被访谈人深刻认识到访谈及其回答的重要性。人们如果认同这件事的重要性会更加配合。不要试图去解决每项抱怨或批评,但告知所有的回答都会得到同等的关注。
3.让自己灵活一些。虽然遵守日程表和提出所有问题十分重要,但少数人可能会在倾听和理解某些问题上遇到麻烦。如果发生了这种情况,放慢速度并重复问题。
4.单独和被访人谈话。其他人在场可能会分散注意力或侵犯隐私,从而改变结果。
5.如果使用印制的访谈调查表,要保证按照卷面的访谈提纲提问。很重要的一点是对所有的人以相同的方式问同样的问

题，否则结果就不具有可比性。如果使用笔记本电脑，你就不必担心问题的顺序。你所要学习的是设备死机或出现其他小故障时如何处理。

6.访谈者应该按照培训班和访谈表上的指示进行访谈。

监控访谈质量。为了确保你获得的是所能得到的最准确的数据，你必须监控访谈的质量。可以采取非正式的形式，比如让访谈者每周给你打个电话。或者采取正式的形式，要求他们向你提交一份他们每日活动的标准化清单。如果可能的话，你可以同访谈员一起去或陪访电话访谈员，以确保他们所做的符合调查的目的。为了避免问题出现，你可能希望采取以下某些或全部步骤：

• 建立一个热线——有人能立即回答可能发生的各种问题，即使是在访谈过程中。

• 为访谈人提供一份书面脚本。如果访谈人要介绍自己或调查，可以给他们一个脚本或一系列可能涉及的主题。

• 确保提供了所有补充材料的额外副本。例如，如果数据收集者是通过邮件把完成的访谈寄回给你，确保提供给他们额外的表格和信封。作为一个可选项，考虑让数据收集者通过电子邮件提供结果或直接录入到一个安全的网站里。

• 准备一份易读的分发资料以描述调查。

• 提供一份日程表或日历，以便访谈者能随时掌握自己的进度。无数软件程序可以帮助你做这件事。

• 考虑给访谈者提供视觉辅助材料。当接受访谈的人因语言能力不足而不能理解调查时，视觉辅助材料特别重要。视觉辅助材料在分类概念和确保每个人是对相同的刺激做出反应方面也是有用的。例如，假设你希望了解人们是否认为经济在发展。为了确保每个人接触的是同一套经济数据，你可以展示从最近一两年的本地报纸中收集的描述经济的图表。视觉辅助材料的另一个用途是在调查人们对一个规划社区的理想期望时。你可以展示几套不同的规划，并要求人们评论每套规划。在访谈中准备音视频辅助材料是相对昂贵的，而且还

需要对访谈人使用这些材料做特别培训。

• 考虑某些访谈人需要重新培训的可能性,并制订相应的计划。

对调查进行测试

你的调查设计好之后,就必须进行测试看它是否能执行以及是否能获得准确的数据。这意味着检验调查的逻辑(访谈者记录回答的容易程度)以及调查表的逻辑。测试的目的是回答以下问题:

• 调查能够提供所需的信息吗?某些词或问题是否是多余的或误导性的?

• 问题适合那些将要接受调查的人们吗?

• 信息收集者能够正确使用调查表吗?他们能够根据书面指导或特别的编码表来展开、收集和报告信息吗?

• 程序是标准化的吗?如果是由人来完成的,能够保证每个人以相同的方式收集信息吗?

• 从调查中获得的信息的一致性如何?

信度和效度:调查的质量

如果每次测量相同的物体(假设物体没有变化)都产生同样的结果,尺子被认为是一种可信的测量工具。如果今天用码尺量你的身高是6英尺1英寸,和两周前的结果一样,那么它就是可信的。

当然,人们会变化。你今天可能比昨天更疲劳、生气和紧张。人们也会因为他们的经验或学到新东西而改变,但是有意义的变化并不容易随机波动。一个可信的调查提供对重要特征的一致的测量,尽管存在背景噪声。它反映了“真实的”分数——一个没有随机误差的分数。

如果一把尺子提供了对身高的准确测量(没有误差),就可以被认为是一种有效的测量工具。但即使尺子测出你今天的身高和六个月前的身高都是6英尺1英寸(意味着它是可信的),它仍然可能是错误

的。如果尺子没有被正确校准就会出现这种情况，你实际的身高可能是5英尺6英寸。

如果你的调查只是问医院管理者每间病房有多少床位，并且至少两次得到同样的回答，那么它就是可信的测量工具。但是如果你这个调查是为了测量医疗服务的质量，那么你所做的就是一个效度有问题的可信的调查。一个有效的调查总是可信的，但一个可信的调查不总是有效的。

调查是可信的吗？

在评论一个已经公开的调查时（以及在评估自己的调查的质量时），需要提出以下三类信度方面的问题：测试—再测、等效和内部一致性。

首先，调查是否具有测试—再测信度？估计信度的一种方式是看是否某人不止一次对调查问题作出相同的回答。通常通过对同一个群体做两次调查，然后计算两次分数的相关系数来计算测试—再测信度。如果结果之间的相关系数较高就可以认为调查是可信的。也就是说，通常第一次持正面（或负面）态度的人，第二次仍然会持正面（或负面）态度。

第二，不同的调查表是否等效？假设如果使用两种不同的调查表评估同一种态度，你必须确保不论采用哪一种调查表人们的打分都相同。例如，你希望使用调查表A作为前测，调查表B作为后测，就需要检查两个表的等效性，以确保两者没有差别。

等效信度可以这样计算：同一天对相同人群使用同一个调查的两个或更多调查表，或者对随机选择出来的两个或更多群体使用不同的调查表。通过比较每种调查表的平均分和标准差，以及计算每种调查表分数的相关系数来确定等效性。如果不同的调查有着几乎相同的平均分和标准差以及高度的相关性，那么它们就具有高的等效信度。等效信度系数应该比较高。检查系数是否尽可能接近极值。

第三，在测量需要测量的特征、态度或质量方面，调查问题的内部一致性如何？为了检验内部一致性，你需要计算一个称之为alpha系

数(或 Cronbach's alpha,这是一个以首次报告这一统计量的人命名的系数)的统计量。Cronbach's alpha 描述了在测量同一属性或维度时问题之间的相互补充性有多好。

许多调查者完全不关注内部一致性,因为他们不打算使用几个问题去测量一种态度或特征。相反,他们感兴趣的是对每个问题的回答。确定你的调查是否需要考虑内部一致性。这里有两个例子:

■ 示例　　**考虑内部一致性**

开展了一项 10 个问题的访谈,以了解病人对医院医疗服务的满意度。高分意味着满意度高,低分意味着满意度低。需要考虑:在多大程度上这 10 道题测量了医院服务满意度的相同维度? ■

■ 示例　　**不考虑内部一致性**

作为一项研究的一部分,对病人开展一个 10 个问题的访谈,了解医院应该如何改进。8 道题询问不同服务的可能改进,如:可以提供的食物类型、医护人员的可得性等。2 个问题询问病人的年龄和受教育水平。因为这个调查关注的是 8 项不同服务的改进意见和提供关于被访者的年龄和教育背景的信息,因而每个问题是相对独立的。 ■

信度多高合适?标准取决于调查的目的。如果是为了做组间比较(如,公司 A 的员工与公司 B 的员工),信度系数在 0.50 及以上是可以接受的。为了做出个人教育或健康需求的决策,就需要 0.90 的系数。

调查是有效的吗?

如果提供的信息是被调查者的知识、态度、价值和行为的真实反映,那么调查就是有效的。以下问题用于了解一个已公开调查的效度:

• **调查是否具有预测效度**?可以通过找到表明调查预测了某人执行某项任务的能力或按某种方式行事的证据来证明调查的效度。例如,一个医学院的入门考试如果准确地预言了某人在医学院的表现,就是具有预测效度的。建立预测效度的一种方式是调查所有需要

进入医学院的学生并将他们的分数与学校表现进行比较。如果两者之间存在高的正相关或负相关，就可以说调查或测量工具具有预测效度。

• **调查是否具有同时效度**？你可以通过与一个知名且被接受的测量进行比较来验证调查的效度。为了建立一个关于数学态度的新调查的同时效度，对同一群体做已验证过的调查，然后比较两者的分数。也可以只对被调查者做新调查，然后将分数与专家对被调查者态度的判断相比较。新调查和基准调查之间高的相关性意味着具有同时效度。切记，只有当基准测量有效时，具有同时效度的研究才是有价值的。

• **调查是否具有内容效度**？如果能证明其问题准确反映了他们希望测量的那些特征或态度，就能说调查是有效的。例如，如果一个关于政治知识的调查，合理地包含了一组人们讨论或阅读政治事务时经常使用的事实、字句和理论，那么就是具有内容效度的。建立内容效度的常用方式是通过参考关于人格、情感和行为的理论，并请教专家你的问题对于你希望调查的态度或品性是否具有代表性。

• **调查是否具有建构效度**？如果能测量诸如敌意或满意度这样的概念，调查就具有建构效度。通过调查那些专家认为表现或未表现出与某概念相关的行为的人，可以实验性地建立建构效度。如果专家判断具有高度敌意或满意度的人，在设计用来测量敌意和满意度的调查中也能取得高分，就可以认为调查具有建构效度。

选择和改编调查

你可以从网上免费获得成千上万（甚至更多）的调查和调查问题。但要记住，它们都是针对一般受众的，因而可能并不适合于你的调查。另外，你也无法知道这些免费问题是否经过试用或分析或者由谁试用或分析过。或许最重要的是，当你引用一个网上调查，你可能被要求签署一份协议，或者被警告将在你的浏览器里安装“cookies”。免费问题的所失可能大于所得。

下面是从一个为了下载某个调查模板而必须签署的真实的服务协议中摘录出来的部分内容。

■ 示例 节选自一个为了使用免费调查问题(名字已变)的真实的网上服务协议

为了运营和提供服务,我们会收集你的某些信息。作为服务的一部分,我们也会自动上传一些关于你的计算机、服务使用和服务表现方面的信息。我们将按照 Extrasoft 在线隐私声明(http://go. extrasoft. com/fwlink/?LinkID=74170)中描述的那样使用和保护这些信息。某些特殊情况下,我们可能会接触或披露有关你的信息,包括你的通讯内容,以(a)符合法律或满足法律要求或司法程序;(b)保护 Extrasoft 或我们的顾客的权利或财产,包括执行调控你的服务使用的协议或政策;或者(c)有充分理由保证这种接触或披露行为对于保护 Extrasoft 员工、顾客或公众的个人安全是必要的。

服务是基于 Extrasoft 为了自身及其顾客的利益而运营的一个私有计算机网络。作为我们保护服务和保护我们的顾客的努力的一部分,Extrasoft 保留迟滞或阻止任何类型电子邮件或与其他服务器交互通信的权利,或者阻止你违反协议。我们采用的技术或其他工具可能会阻碍或中断你对服务的使用。 ■

关于在研究文献中寻找可用的和有用的调查的指南

确保你的调查可信且可靠的方式之一,是依托一个已经在科学研究中得到其他调查者和研究者仔细检验的调查。如果你希望调查态度、情感、健康状况、生活质量和道德价值观,寻找科学检验过的测量尤为重要。这些因素(以及其他类似的),是含义模糊且难以测量的,并且建立有效测量需要专业的研究知识和资金帮助。如果你并不打算做这样一个专业研究,那么就需要找到一个可信有效且合适你需求的调查。

你可以通过已发表的文献寻找调查,或者利用民意调查网站或其他保存公众价值、健康、教育和福利数据的网站。

许多健康、教育和社会科学领域的研究者开发、验证和使用调查方法,并把他们的结果发表在同行评议的杂志上。已发表文献是搜寻

已有调查问卷和调查问题的极佳渠道。但是你怎样才能在全世界海量且不断增长的文献中挖掘出这些调查呢？并且，就算你找到它们，如何从中做出选择？可以根据下列问题了解如何搜索文献并评估其质量和适当性。

1.**选择一个或多个与你的调查相关的网上数字图书馆**。网上图书馆包含大量世界上已发表的研究文章。笔者经常使用的网上图书馆包括 PubMed（医疗和健康）、ERIC（教育）和 Web of Science（社会科学和科学）。这些数字图书馆提供某些杂志上已发表文章的免费接入。大多数大学和某些商业和公共图书馆订购了一个或多个网上文献数据库，通过它们，学生、员工和研究者可以阅读世界范围的杂志。The Public Library of Science（PLoS）有一个可供所有人免费使用的杂志公开接口。Google Scholar 也提供包含调查的杂志文章的接入。Google Scholar 不是为学术和调查目的而专门设计的，因而找到你需要的东西所花费的时间可能长于 Web of Science 这样的数据库。

2.**明确你的调查希望涉及的主题或变量**。比较以下两个情境：

调查者 1 希望找出为什么一个新的课后项目的学生会退出。调查者 1 制定了一个课后项目退出调查。这个调查提出的问题包括家庭收入、家庭孩子数以及所有家庭成员的教育水平。与那些留在项目中的学生相比，调查者 1 假设退出的学生更可能来自较低收入和教育水平的家庭。

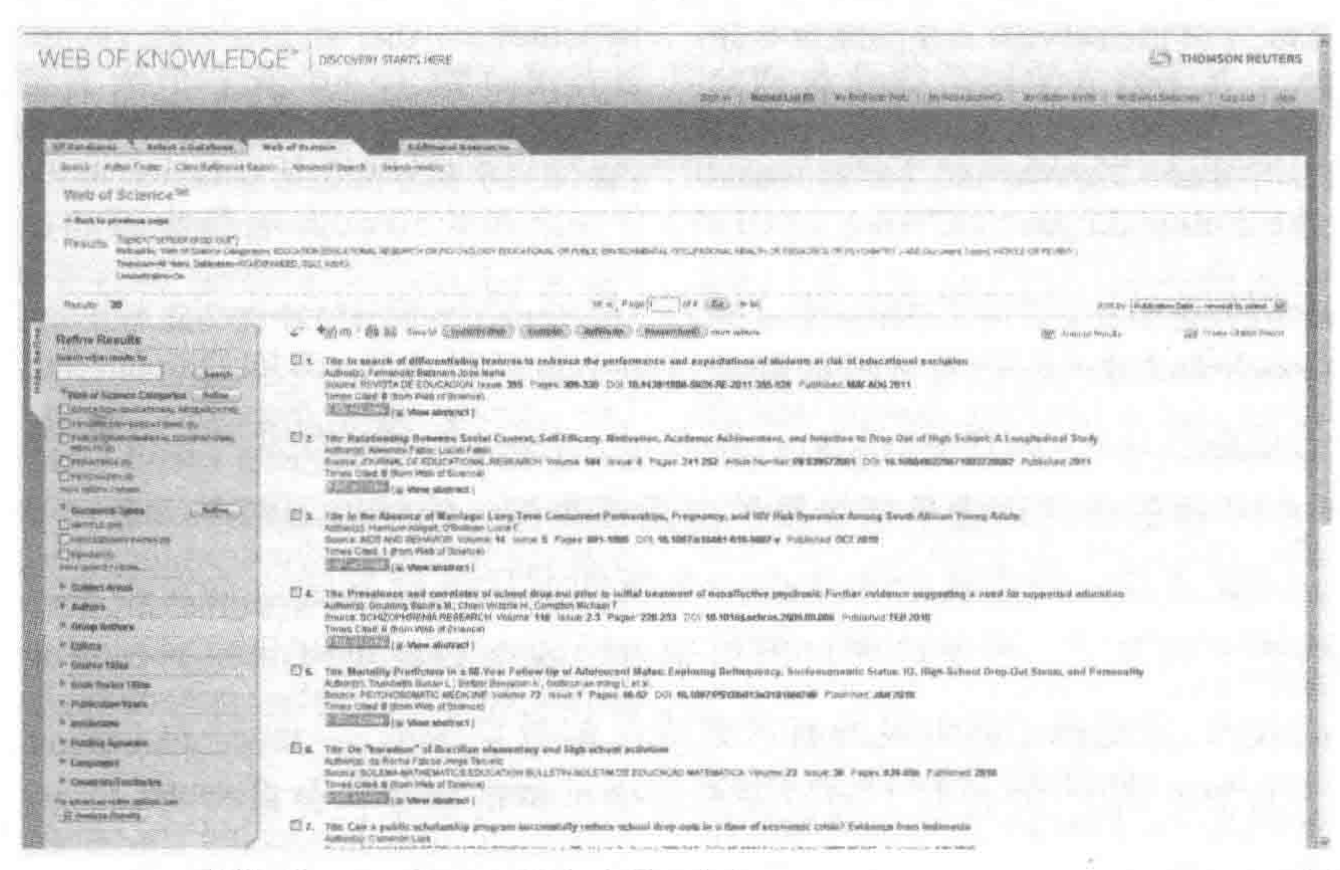
调查者 2 也对找出新的课后项目学生退出的原因感兴趣。调查者 2 制定的调查提出的问题包括学生每天用于兴趣阅读、玩视频游戏或网上访问朋友的时间。调查者 2 假设花费多数时间在玩视频游戏或上网的学生，比那些花费大部分时间在兴趣阅读上的学生更容易退出项目。

如果你对研究家庭收入和教育是否与留在课后项目相关感兴趣，那么调查 1 是一个好的选择。如果你对退出课后项目与学生花费在视频游戏和 Facebook 上的时间之间的关系感兴趣，那么调查 2 是适合你的。注意调查者 1 和调查者 2 都给他们的调查取名叫“课后项目退出调查”。具有相同或类似名称的调查可能实际测量的是非常不同的

概念。你需要事先明确你希望了解什么,然后仔细评估每项可得调查的适当性。

3.**学会如何寻找合适的电子图书馆**。所有的电子图书馆都为搜索他们的数据库提供指导,并且相对使用容易。这里有一个 Web of Science 上的辍学调查的检索结果的例子。你可以通过选择学科、文章类型、语言、出版年等来进一步精炼你的检索。

■示例　　**检索辍学调查:来自 Web of Science 的示例**



来源:Thomson Reuters Web of Knowledge. ■

一旦你获得了潜在可用文章的列表,逐篇检查是否包含调查。如果你因为没有找到合适的调查而需要进一步检索,可以研究一下相关文章的关键词。你经常可以在文章摘要或概要旁边找到关键词。每个电子图书馆都使用主题词和描述符来分类其文章。你可以从图书馆的教程里学习如何使用它们。

4.**确信可以按期望的方式开展调查或有足够的资源改编调查**。假设你在相关主题的文献中找到一个调查,它研究的主题和具体问题都与你的相似,但是这个调查采用的是面对面访谈,而你希望的是网上调查。你需要相当多的专业知识把访谈问卷转化为一个网上调查。

一旦完成,需要做大量小规模测试。转化、小规模测试和基于测试的调整会延迟调查开始的时间,而且是昂贵的。所以,即使这个调查看上去不错,对你当前的需求而言可能也是不合适的。你还得再多看看。

5.**检查被调查者的特征**。假设你找到一篇文章,里面有个看上去很不错的你可以借鉴的调查。当你阅读使用这个调查的这篇文章时,问一下自己,它的被调查者与你计划中的样本是否相似。例如,一个设计用来了解孩子如何应对暴力的调查与一个研究成年人的调查,使用的语言和概念是有很大不同的。

6.**选择具有最佳证据证明其信度和效度良好的调查**。最佳证据来自于研究方法的严谨。这些研究被仔细对照(如,随机对照试验),具有合理和有意义的样本、适当的统计和好的结论。它们描述和论证了所有他们所用测量(包括调查)的选择理由。到参考文献列出的原始来源去检查调查是否经过验证,这往往是个好主意。

以下给出三个研究文献关于信度和效度内容的例子。

■ 示例　　**研究文献中的信度和效度的示例性报告**

测量恋爱关系中的爱:一个元分析

这项研究(Graham, 2011)的目的在于识别目前构成当前常用的大多数关于爱的测量的基础高级因素。研究者使用了一种其称为"元分析因素分析"的方法。81 项研究涉及的 103 个样本和 19 387 个人为元分析提供了数据。几种常用恋爱测量的报告相关系数被汇总到一个元分析相关系数矩阵中,然后对该矩阵进行主成分分析。研究显示,存在着一般性的爱、迷恋和实际友谊等因素。爱与关系满意度和持续时间呈现正相关,而迷恋则呈负相关。这个结果确认了最近对爱的概念化(同时效度),并且指出了恋爱研究领域中对除整体自陈以外的其他测量方法的需求。

女青少年犯的精神病态评定量表(年轻版)的信度和建构效度

虽然最近的研究已经表明评估男性青少年的精神病态特征的量表的有用性,但对于评估女性青少年的精神病态特征的量表的有用性的证据则少很多。这项研究(Bauer, Whitman, & Kosson, 2011)利用一个 80 名女青少年犯的样本,考察了精神病态评定表(年轻版)(PCL:YV)的信度和建构效

度。信度分析显示了 PCL:YV 分数具有较高的评分者一致性和内部一致性。与关于成年和男性青少年犯的研究相一致,精神病态特征与下列特征相关:大量品行障碍和酒精依赖症状、较早出现的犯罪行为、违反制度的癖好和与精神病态相关的人际行为(建构效度)。与一些早前的青少年研究相一致,PCL:YV 分数与内化性障碍诊断之间不存在负相关关系,而在精神病态评级和消极情感之间则存在某些正相关关系(建构效度)。结果提示评估女性青少年的精神病态特征,为青少年犯罪行为、模式和个人特征提供了有意义的信息。

快餐消费的决定因素:计划行为理论的一个应用

这项研究(Dunn, Mohr, Wilson, & Wittert, 2011)考察了影响澳大利亚快餐消费的因素。研究者应用计划行为理论(内容效度)调查了 404 人,了解他们的态度、意愿和行为。为了研究预测快餐消费的因素(预测效度),研究者使用了结构方程模型的技术。他们还使用其他方法,包括因子分析的方法,提炼由对快餐的态度构成的概念的定义。研究者发现快餐消费会受到方便性、满意度、家庭和朋友、快节奏生活方式、对快餐强烈欲望的体验、不会做饭、工作时间长和独自吃饭等因素的影响。人们对长期健康的关注可能被快餐带来的短期利益压倒。 ■

7.**对一小组人试用调查,以观察其可用性和有用性**。然后再扩大测试规模以确保万无一失。

在网上寻找调查

有些组织在网上向公众免费提供调查问题。以下只是少数例子。

盖洛普公司(http://www.gallup.com)

在这个网站上,你可以找到最近的政治和经济方面的投票问题及结果,以及对社会问题和消费者行为的态度的"特别报告"。

一般社会调查(http://www.norc.org/Research/Projects/Pages/general-social-survey.aspx)

全国民意研究中心自 1972 年以来一直开展一般社会调查(the General Social Survey)。它测量有关社会和公共政策、经济状况、政治事件、工作和家庭生活方面的态度。

全国民意研究中心(http://norc.uchicago.edu/homepage.htm)

NORC 为许多机构和公司开展公共政策方面的调查研究。在“项目”一栏可以找到他们最近的调查和报告。

皮尤慈善信托基金(http://www.pewtrusts.org)

皮尤基金提供以下方面的调查和调查问题:大众传媒、宗教和公共生活、社会和互联网和全球态度以及拉丁裔美国人和公众意见。

美国医疗保健研究与质量机构(AHRQ)(http://www.ahrq.gov/data/dataresources.htm)

在这个网站,你能找到可以帮助你开展有关卫生保健差距、质量和安全性以及艾滋病等方面调查的信息。

美国人口普查(http://www.census.gov/)

这是调查问题的一个主要来源。无论调查是大还是小,甚或在美国之外开展,在设计和改编有关年龄、民族、收入、教育和职业方面的调查问题上,这都是一个极好的网站。美国人口普查还提供与刑事审判相关的调查问题(http://www.census.gov/govs/cj)。有包括年度假释和缓刑调查和性暴力调查在内的随机调查。

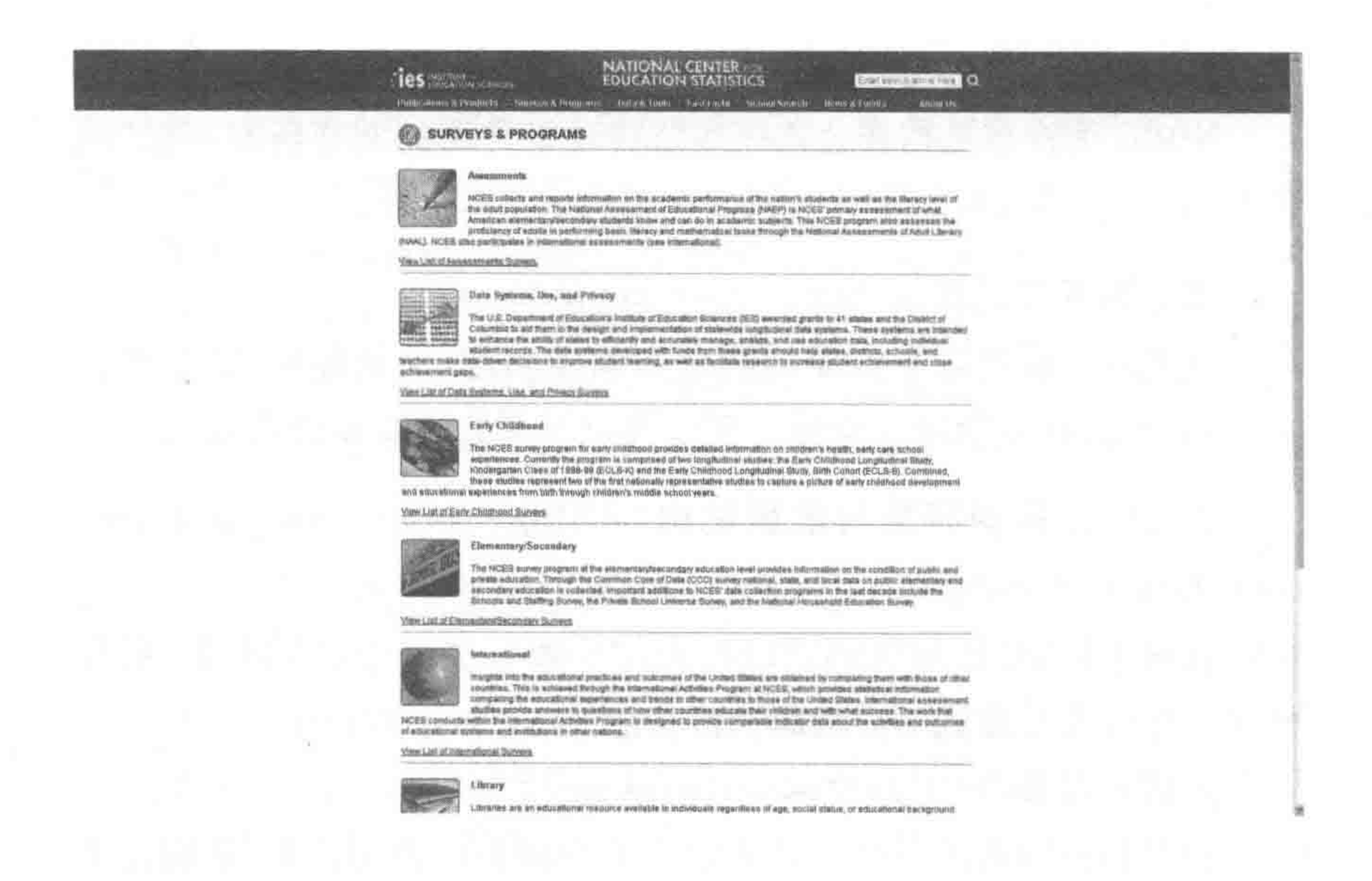
全国教育统计中心里的教育科学研究所(http://nces.ed.gov/datatools/index.asp? DataToolSectionID=3)

这个网站是个调查大观园。以下是你登录该网站后能够找到的调查的例子。正如你所见,还包括国际调查。

■ 示例 **搜索教育调查:来自全国教育统计中心的一个示例页面(见下页)**

■

记住,买主,要小心! 仔细选择那些与你的被调查者相似的调查和个别问题。有些网上调查公司提供从顾客满意度到健康状况的广阔领域内的免费调查问题。这些问题有很多从来没有被任何人检验或使用过,因而它们的价值不明。避免不良调查和调查问题的最佳方式是进行小规模测试或试用。



小规模测试的指南

以下是一个好的调查测试的一些基本原则:

1.预想调查开展的真实环境并制订相应预案。对于访谈,这意味着重新制定培训手册和所有的表格;对于邮寄调查问卷,你不得不制作封面信、回邮信封等。对于网上调查,你需要检验被调查者登录一个安全网站以及进入调查的容易程度。不用说,这需要计划和时间,可能花费不菲,因此要确保你有开展这些活动的充足预算。

2.你可以以非正式方式试用调查选定的部分来开始。例如,先对自填问卷的导语部分进行测试,或者对访谈中的几个问题的措辞进行测试。你还可以采用你最终可能使用的几种不同方法先测试调查程序。所以如果你计划向参会人员分发调查问卷,测验必须包含一个访谈,讨论与调查表中的问题相关的所有问题。最后,你需要对调查的安排和形式进行预先测试。

3.选择与你最终完成调查相似的被调查者。他们应该年龄相仿、教育水平相近等。

4.在不浪费资源的情况下招募合理数量的人参与测试。或许测

试一个 5 题的问卷比 20 题的问卷所需的人数要少。另外,如果你发现对调查所需的改进不大了,那么就停止。

5.对于信度,集中在问题的清晰性和调查的一般形式上。以下是可能遇到的情况:

- 未能回答问题
- 对同一个问题给出几个答案
- 在书面调查的页面边缘书写评语

出现上述任何问题是问卷不可信并需要修改的信号。选择题的选项是否互斥?是否提供了所有可能的选项?问卷或访谈的语言是否清楚和无歧义?导语和过渡语是否顺畅?是否选择了正确的问题顺序?问卷是否过长或难以阅读?访谈是否过长?(例如,你计划开展一个 10 分钟的访谈,但是你在预先测试时花费了 20 分钟。)

为了提高效度,你必须确信所有相关的主题都被包含在调查里(在给定的资源条件下)。比如一个政治态度的调查,你是否包含了所有党派?争议性问题?为了使你的调查具有内容效度还需纳入哪些问题?如果你不确定,和其他人一起检查,包括试调查。调查问卷是否留有足够空白表达所有的观点?假设你在开展一项关于人们的宗教虔诚的调查,如果事先有证据表明他们都十分虔诚,你可能不需要再调查了。除非你能显示至少在理论上可以区分虔诚和不虔诚的信徒,否则没有人相信这个结果。你如何弥补这一点?在测试中,选择那些你知道他们虔诚的人和你知道他们不虔诚的人进行调查,并确认他们的反应是否真的不同?

检验回答范围的广度。如果人们的观点或感觉确实不同,调查能否捕捉到这些差异?假设你调查一个郊区邻里社区对新建一座高层建筑的态度。你需要同时调查支持和反对的人。这种方法能帮助揭示你自己在问题的措辞方面的偏见,对于封闭问题,还能帮助找出你之前没有想到的认为不存在的人们的强烈感受。

考虑这一点:在一个儿童自我报告健康行为的调查的测试中,被调查者被提问餐后洗手的频度。所有参与调查的 6 个年龄在 8 到 10 岁的儿童,在"总是""从不"和"我不知道"的选项中做出了"总是"的

回答。选项被变更为“几乎总是”“经常”和“几乎从不”。在新的分类中,同样的6个孩子有2个回答“几乎总是”,4个回答“经常”。

更进一步:调查、语言和文化

许多调查被翻译为不同的语言。如果你打算翻译你的调查,不要指望翻译软件可以自动进行,仅仅翻译也是不够的。在原来的语言和另外的语言之间横亘着一条文化鸿沟,你可能需要针对每个调查问题进行改写。

为了避免因为误解他们的语言或文化而弄晕或惹恼人们,你需要遵循一些简单的原则。这些原则包括招募语言(外语或方言)流利的助手,以及对典型的被调查者进行试调查。遵守以下指导原则:

[翻译调查问卷的指导原则]

- 使用语言流利的人做第一道翻译。如果可以的话,使用母语译者。如果你能够负担,那就找专业译者。翻译的艺术是十分微妙的——单词和词组需要多年的文化浸淫才能掌握。如果你使用流利的译者,就可以减少修改问题的措辞和回答选项的时间。
- 给3~5名母语人士试用翻译好的调查问卷。向他们提出如下问题:这个问题问的是什么?你能想出更好的提问方式吗?
- 在原来译者的帮助下修改调查问卷。
- 把调查问卷翻译回原来的语言。使用另外一位译者来完成这个任务。这个“回译”的调查问卷是否与原来的版本相吻合?如果不吻合的话,两位译者应该合作使其吻合。
- 接下来把问卷交给目标被调查者的一个小组(5~10人)试用。
- 修改调查问卷。
- 小规模测试调查问卷。

- 生成最终版本。
- 如果你希望知道被调查者的背景或种族,你可能最好使用现有的问题而不是自创。以下关于种族的问题来自美国健康和人类服务部(http:/minorityhealth. hhs. gov/templates/content.aspx? ID=9227&1v1ID=208)。

■示例 **关于种族的问题:美国健康和人类服务部少数民族健康办公室**

你是拉美裔、拉丁美洲人或西班牙血统吗?(可以选择一项或多项)

a)不是,不是拉美裔、拉丁美洲人或西班牙血统

b)是的,墨西哥人,墨西哥美国人,奇卡诺人

c)是的,波多黎各人

d)是的,古巴人

e)是的,其他拉美裔、拉丁美洲人或西班牙血统

你的人种是什么?(可以选择一项或多项)

a)白人

b)黑人或非洲裔美国人

c)美国印第安人或阿拉斯加土著

d)亚洲印度人

e)中国人

f)菲律宾人

g)日本人

h)韩国人

i)越南人

j)其他亚洲人

k)夏威夷土著

l)关岛人或夏莫洛人

m)萨摩亚岛人

n)其他太平洋岛屿人

这些关于种族和民族的问题是与美国人口相关的。在英国,使用其他分类,如白人、亚洲人(印度人、巴基斯坦人、孟加拉人)、黑人(加勒比海人、非洲人)、混合人种和中国人(http://www.communities.gov.uk/publications/

corporate/statistics/citizenshipsurvey200809equality)。 ■

小 结

调查问卷的长度由你需要了解什么,以及需要提出多少问题才能使相应的结果可信来决定。另外的一个考虑因素是被调查者能拿出多少时间和注意力用于调查?

问题应该从最熟悉的到最不熟悉的进行排列。

把相对容易回答的问题放在调查的末尾。避免许多相似的问题。

避免在书面自填问卷中使用跳转模式。网上调查是处理分支问题的有效工具,因为你可以设计软件指导被调查者,并且每家商业软件供应商都提供跳转模式的功能。

好的电话访谈或面对面访谈的关键是培训。培训能确保让所有的访谈人了解对他们的期望,让他们在相同长短的时间内以相同的方式提出所有问题。

应该对访谈予以监控,以确保访谈员以相同的方式提出问题并使回答的一致性最大化(信度)。

小规模测试能够提高信度和效度,因为它能帮助你确定问题是否包含了所有主题,是否提供了充分的回答选项。

可信的调查提供一致的信息。

有效的调查提供准确的信息。

确保你的调查可信且有效的一个方式是使用一个已经通过仔细的测试被证明为可信且有效的测量。

测试—再测信度通常是这样计算得出的:对同一群体在两个时点分别调查,然后计算两个时点分数的相关系数。

等效信度可以这样计算:同一天对相同人群使用同一个调查的两个或更多调查表,或者对随机选择出来的两个或更多群体使用不同的调查表。

另外一个信度度量是在测量需要测量的特征、态度或质量的调查中,调查问题的内部一致性如何。为了检验内部一致性,需要计算一

个称作 alpha 系数(或 Cronbach's alpha)的统计量。

预测效度指的是调查的数据预测被调查者执行某项任务的能力或按某种方式行事的准确程度。

同时效度指的是来自一个新调查的数据与来自之前调查的数据相关。

如果专家认为其问题或题项准确地反映了希望测量的特征或态度,那么调查就是具有内容效度的。

当一项关于某个概念的调查,被实践经验丰富的专家认可,在试调查中人们与这个概念相关或无关的行为确实被测到时,就具有了建构效度。

如果你打算翻译调查,招募语言(含外语和方言)流利的人帮助你,并对典型的被调查者进行试调查。

仔细选择已发表的调查。检查它们是否可信、有效且适合你的调查对象。

思考一下

1.找到后面所列的文章并回答下面的问题:

a)命名文章中讨论的调查。

b)描述每个调查的目的。

c)讨论每个调查的信度和效度。

d)你认为调查的信度和效度令人满意吗?请加以解释。

2.任意选择 3 个你感兴趣的调查并回答问题 1 中的几个问题。

3.阅读下面的案例研究。第三位教授获取调查信息的行为,即使他负责向被调查者邮寄补贴支票,是否妥当?

三位教授从全国教育图书馆得到一笔资金,对 60 所学校的数据交换系统进行一个网络调查。每位教授执教于不同的大学。两位教授获得了各自伦理委员会的允许,可以获取被调查者的电子邮件地址以发出调查、跟踪未回应者和开展统计分析。第三位教授没有获得明确的许可做这些事情,但是他负责向调查参与者发放补贴。三位教授使用 E-line Online Survey 公司的服务建立了调查,如果有正确的用户名和密码,任何人(包括第三位教授)都可以登录调查网站,并可以分人、分问题地查看结果。第三位教授需要获得电子邮件地址以便索要邮寄地址。因为没有获得伦理委员会的允许,所以他没有查看调查的任何回答。他向被调查者发送电子邮件,然后获取他们的邮寄地址,再寄出补贴支票。

文 献

Asgeirsdottir, B.B., Sigfusdottir, I.D., Gudljonsson, G.H., & Sigurdsson, J.F.(2011). Associations between sexual abuse and family conflict/violence, self-injurious behavior, and substance use: The mediating role of depressed mood and anger. *Child Abuse & Neglect*, 35(3), 210-219.

Bauer, D.L., Whitman, L.A., & Kosson, D.S.(2011). Reliability and construct validity of psychopathy checklist. *Criminal Justice and Behavior*, 38(10), 965-987.

Dunn, K.I., Mohr, P., Wilson, C.J., & Wittert, G.A.(2011). Determinants of fast-food consumption. An application of the theory of planned behaviour. *Appetite*, 57(2), 349-357. doi:10.1016/j.appet.2011.06.004.

Flynn, K.E., Smith, M.A., & Freese, J.(2006). When do older adults turn to the Internet for health information? Findings from the Wisconsin Longitudinal Study. *Journal of General Internal Medicine*, 21(12), 1295-1301.

Graham, J.M.(2011). Measuring love in romantic relationships: A meta-analysis. *Journal of Social and Personal Relationships*, 28(6), 748-771.

Holman, E.A., Silver, R.C., Poulin, M., Andersen, J., Gil-Rivas, V., & McIntosh, D.N. (2008). Terrorism, acute stress, and cardiovascular health: A 3-year national study following the September 11th attacks. *Archives of General Psychiatry*, 65(1), 73-80.

O'Connor, T.M., Yang, S.-J., & Nicklas, T.A. (2006). Beverage intake among preschool children and its effect on weight status. *Pediatrics*, 118(4), e101()-e1018.

Overby, M., Carrell, T., & Bernthal, J. (2007). Teachers' perceptions of students with speech sound disorders: A quantitative and qualitative analysis. *Language, Speech, & Hearing Services in Schools*, 38(4), 327-341.

Richardson, J., Coid, J., Petruckevitch, A., Chung, W. S., Moorey, S., & Feder, G. (2002). Identifying domestic violence: Cross sectional study in primary care. *British Medical Journal*, 324(7332), 274.

Skoffer, B. (2007). Low back pain in15-to16-year-old children in relation to school furniture and carrying of the school Bag. *Spine*, 32(24), E713.

Wolfson, M., Champion, H., Rogers, T., Neiberg, R.H., Barker, D.C., Talton, J.W., et al. (2011). Evaluation of free to grow: Head start partnerships to promote substance-free communities. *Evaluation Review*. 35(2), 153-188.

4

调查抽样



你需要调查所有人还是只需要抽取一个样本？答案取决于你需要多快取得结果，如果遗漏了某些人或群体是否会损害研究发现的可信度，以及你的资金和技术资源。

抽样可以划分为两种类型。随机抽样是通过客观的方法（诸如从帽子里随机抽取名字）抽样，你可以计算每个人被抽中的概率。非随机抽样是便利的：你只选择那些就近的和愿意并且能够完成调查的人。

样本多大合适？相对较大的样本会减少抽样误差。如果你想评估两个（或更多）群体的表现，统计方法能帮助你选择能够检出组间差异（如果差异真实存在的话）的足够大的样本量。统计方法还能根据你的样本量告诉你，你能观察到的差异有多大。

回答率是将完成调查的人数除以符合条件的被调查者人数。为了提高回答率，确保被调查者能够容易地完成和提交调查，要培训调查员开展调查并跟踪、监控调查执行的质量，为回答保密，以及在可能和合乎道德的前提下向被调查者提供奖金和回报。

样本量和回答率:谁和多少?

当你开展一个调查时,你是否纳入了所有人?答案可能为否。通常的情况是,目标群体的"所有人"等于"很多很多的人",全部调查他们是不现实的。

如果你决定抽样,你就必须提出下列问题:应该纳入多少人?如果你的公司有 1000 名员工,而你决定只调查其中的部分人,你如何决定调查多少人?假设你是想比较长期和短期的工作满意度。统计方法能帮助你确定每个时间(长期和短期)需要纳入多少人,才能在不同时期确实存在差异的时候,检测出这种差异。比如说你有 3 个月时间开展调查,你认为你可以调查 300 位公司员工。统计方法能够帮助你确定,在你的给定资源能够负担的样本量之下,能够检出多大的效应。

假设你想了解你的邻居是否支持一个社区守望项目。该项目中,每个家庭至少守望另外一个家庭(当其屋主外出时)。再假定你的社区有 1 000 户。你是否需要调查所有住户?如果是的话,这样做是否会对项目更有利?以下是一些你需要回答的问题。

1.**数据需求的紧迫程度如何**?假设最近入户抢劫案件增加,因而开展调查并希望尽快启动。如果要调查完邻里所有的 1 000 户,就浪费了宝贵的时间。

2.**计划做什么类型的调查**?如果你准备采用电话或网上调查,调查花费的时间可能少于入户访谈或等待人们寄回调查问卷。

3.**你的资源如何**?如果资源有限,你最好选择诸如电话访谈这样的方法,而不是入户访谈。

4.**你的研究发现的可信度如何**?如果调查所有的 1 000 户,无疑可以认为你的调查代表了整个邻里。如果只有 10 户参加,你就会遇到可信度问题。

5.**你对抽样方法的熟悉程度如何**?抽样方法可以简单可以复杂。全国民意测验和大型学术研究采用高级的技术,需要统计学家和其他专家的专业支持才行。

你如何选择一个样本？你可以使用随机抽样或方便抽样方法。随机抽样是一种客观的选择。一个例子是使用特殊软件随机选择获奖彩票的号码。有统计方法可以计算每个人被选中（或彩票获奖）的概率。方便抽样是选择找得到的并且愿意参加调查的人。方便抽样选择程序不能被认为是“客观的”，因为不是每个符合条件的人都有同等的机会。如果你在调查的当天病了，那么即使你满足其他所有要求也不能参加。

考虑下面两个案例。

■ 示例　　随机抽样和方便抽样

案例 1. 一项调查的目的是了解 Loeb 学校的老师对某些学校改革措施的感觉。所有 100 名教师的名字被放入一顶帽子中，搅乱后由校长选择 50 人参与调查。

案例 2. 开展一项调查了解 Los Hadas 学区的教师对某些学校改革措施的意见。从该学区的 6 所小学中选择 10 名教师进行访谈，从 4 所中学中选择 15 名，从 1 所高级中学中选择 60 名。参与的教师是由各自校长推荐的志愿者。他们满足下列条件：已经在学区执教 5 年或以上，属于三个教师协会或联盟之一，在过去的一年里至少参加过一次关于学区学校改革的会议。

■

在第一个调查中，从帽子中的 100 名教师中抽取 50 人的样本。这种类型的抽样符合随机抽样，因为选择方法是客观的，你可以使用数学公式计算每个人被选中的概率。另外，你通常可以预期两组教师之间不会存在系统性的差异。在每组的 50 人中，你可能有相似数量的聪明、健康和大度的人，也可能有同样数量的不聪明、不健康、不大度的人。

在第二个调查中，校长做出推荐，然后满足条件的教师志愿选择参与。校长可能有他们的偏好。志愿教师可能系统性地不同于非志愿者。例如，他们对调查可能更热情，或者有更多时间完成调查。这种抽样策略不是客观的，它属于非随机抽样。

随机抽样方法

以下是三种最常采用的随机抽样方法:

- 简单随机抽样
- 分层随机抽样
- 简单随机整群抽样
- 分层随机整群抽样*

简单随机抽样是总体中每个个体具有同等机会被选中的抽样方法。总体包含符合调查要求的所有人。

下面是一个简单随机抽样。

■ 示例　　简单随机抽样

你希望从慈善基金会中选择 100 人调查他们的资助类型。总共有 400 人可以提供此类信息。你把他们的名字按顺序排好。每个名字赋予一个从 001 到 400 的数字。然后,使用随机数表(在统计教科书中可以找到),选择数字最先出现的头 100 人。 ■

在这个场景中每个人都有同样的机会被选中。总体包含 400 人。谁会被选中只取决于运气。

以下不是随机抽样。

■ 示例　　非随机抽样

你希望从慈善基金会中选择 100 人调查他们的资助类型。总共有 400 人可以提供此类信息。你从全国的 4 个区域中各选择 25 人。 ■

在这个场景中有些人没有被选中的机会:那些不住在你指定的 4 个地理区域中的人。

下面是另外一个简单随机抽样的例子。

---

* 原文有误,似乎不应该有这项。——译者注

■ 示例　　　　**简单随机抽样**

一个中西部城市的200名护士、药剂师和社会工作者报名参加一个成人日间照料研讨班。该市只能负担50人的费用。研讨班主任使用统计程序提供的随机分配功能选出了50个人。 ■

针对电话调查的随机抽样的一些特殊情况，一些调查者使用一种被称为随机数字拨号的技术。加一法是其变种之一，也就是在实际选中的电话号码上加一个数字。如果选中的电话号码是(311)459-4231，则拨打电话号码(311)459-4232。这一技术解决了在这个国家的很多地区，特别是城市地区，人们没有登记他们的电话号码的问题。这些人不应该不公平地被电话调查遗漏。

做出决定

你何时需要选择简单随机抽样？简单随机抽样的优点如下：

- 它是所有随机抽样方法中最简单的。
- 能够获取帮助。大多数统计教科书都有易于使用的表格供随机抽样用。另外一种方法是使用所有统计软件都提供的随机数功能。

简单随机抽样的一个主要缺点是其不能被用来对被调查者分群或层(如，60%的男性和40%的女性)。为了确保样本的比例符合你的需求，你需要分层。

分层随机抽样

在简单随机抽样中，你可以从一个总体中随机地选取一组被访者。在分层随机抽样中，你首先需要把总体细分成亚群或层，然后从每层中选择一定数量或比例的被访者以获得一个样本。

举例来说，你可以使用分层随机抽样方法来获得对男性和女性相同的代表性。你是通过把整群划分成男性和女性的亚群，然后从每个亚群中随机选取一定数量的被访者来做到这一点的。这种抽样方法比简单随机抽样更精确，因为它使群同质化，不过仅在正确地选择群时才是这样。也就是说，不要分别抽取男性和女性，除非你准备在他

们之间进行比较。只有事先确信这种比较是有意义的时候才进行这种比较。例如,在一个投票人偏好的调查中,如果你有某些证据表明男性和女性的投票不同,此时才需要保证你的调查中包含足够多的男性和女性以进行比较。如果只使用简单随机抽样,你可能发现你的调查样本由于偶然因素而主要是由男性或主要由女性组成的。

下面是如何进行分层随机抽样的例子。

■ 示例 **分层随机抽样**

大学健康中心正在考虑开展一个新项目以帮助年轻人减肥。在改变项目前,管理部门打算开展一个调查,以了解新项目与老项目相比如何,以及不同年龄的男女学生的表现。之前的经验显示年纪较大的学生在减肥项目中成效更大。所以,调查者计划抽取一个由两个年龄群——17 到 22 岁和 23 到 28 岁——的男性和女性组成的样本,然后比较他们在每个项目中的表现。

大约 310 名大学生报名参加健康中心冬季学期的常规减肥项目。在这 310 名参加者中,140 人在 17 到 22 岁之间,其中 62 人是男性。170 人在 23 到 28 岁之间,其中 80 人是男性。调查者从这四个子群(男性、女性、年龄 17~22岁以及年龄 23~28 岁)中分别随机选出 40 人,然后每隔一人指派一个学生到新项目。样本如下:

大学健康中心的减肥项目

	年龄 17~22 岁		年龄 23~28 岁		
	男　性	女　性	男　性	女　性	合　计
常规项目	20	20	20	20	80
新项目	20	20	20	20	80
总计	40	40	40	40	160

■

做出决定

分层随机抽样的一个优点是调查者能够选择一个代表目标总体各种群体和特征模式的样本。

分层随机抽样的缺点是它需要比简单随机抽样付出更多努力,并

且为了产生具有统计意义的结果，它经常需要一个比简单随机抽样更大的样本量。记住，对于每一层或子群体，你可能需要 20~30 人才能做出有意义的统计比较。

如果你在选用分层随机抽样方面有困难，记住通过分层增加的精度一般也可以通过增加简单随机抽样的样本量来获得。增加样本量可能比实施分层随机抽样更容易。

简单随机整群抽样

使用简单随机整群抽样主要是出于管理上的便利，而不是为了增加抽样精度。有时候随机选择个人是不行的：例如，如果每个病房只选择少数病人调查的话，就会打搅到所有的病房。有时候个体的随机选择在管理上不可能的。

解决个体作为抽样单位带来的问题的一个方法是使用被调查者的群体或群集。

在简单随机抽样中，你从所有参与调查的可能个体中随机选择一个子集。整群抽样与随机抽样类似，除了随机指派的是群体而不是个体以外。这种方法假定总体是由自然的或预设的群集或群体组成。以下是其原理。

■ 示例　　**简单随机整群抽样**

社区精神健康中心有 40 个独立的家庭辅导组，每组有大约 30 个成员。中心主任注意到去年出席率出现下降，因而决定尝试开展一个实验性项目。在该项目中，每个人将在开始治疗前接受单独的检查和问询。这个项目是昂贵的，中心开始的时候只能负担 150 人的项目费用。

从所有群体成员中随机选择个人将会带来抵触和某些群体完整性的破坏。作为替代，采用了简单随机整群抽样的方案，随机选择了 5 个 30 人的群体——总共 150 人——参与实验性项目。每个群体被作为一个群集对待。在 6 个月后，将实验性项目的效果与传统项目进行了比较。 ■

做出决定

简单随机整群抽样的优点是：

- 它可以用于随机抽选个体存在不方便或不道德的情形。
- 它简化了调查管理。

简单随机整群抽样的缺点是它需要复杂的统计方法去使抽样单位(医院、街道、学校)和分析单位(病人、家务工作者、学生)相吻合。也就是说,你根据群集(诸如学校)抽样,但却是在个体(诸如学生)层面分析数据。

虽然在例子中有150人参与调查,但实际上只有5个单位(5个30人的群体)需要研究。为什么你不能分别研究这150人中的每一个人?当人们处于特殊的群体中——班级、俱乐部、组织、邻里——他们倾向于共享相似的特征和观点。这种情况下研究每个人可能是多余的,因为某人可能与另外一个人相似。你需要每个群集有相对数量较多的人以使这个问题不那么显著。具体需要多少人是一个统计学问题。

复杂的抽样策略需要了解抽样统计的相关知识。

系统抽样

在系统抽样中,你选择一个数字,比如说5,然后在总体名单中每隔5个名字选择一个。如果名单包含10 000个名字,而调查者需要一个1 000人的样本,那么可以每隔10个名字选一个。

假设你有一个500个名字的名单,你希望从中选出100人。你可以随机选择一个1到10之间的数字。如果你选择数字3,你可以从名单中的第3个名字开始每5个名字选择一个。你的样本选择将会是第3个名字、第8个、第13个等,直到选够了100个名字。如果你随机选择"开始"数字,那么系统抽样很像简单随机抽样。

做出决定

系统抽样中存在风险。人名单中可能隐含着某种未被察觉的模式。如果你使用这样的名单,你的样本就很容易因为这种模式而产生偏倚。例如,在美国相对很少有人的姓以W、X、Y、Z开头,因而可能在

系统抽样中被低估。这里是另外一个例子。假设你抽样教室以调查学生对学校的态度。再比如教室是按下面的顺序排列的：

1层	1a	1b	1c
2层	2a	2b	2c
*N*层	*N*a	*N*b	*N*c

进一步假设你从1a开始每3个教室选择1个。样本将包含教室1a、4a、7a等，直到*N*a。如果所有"a"的教室都是面向草坪因而安静的，而"b"和"c"的教室面向运动场因而喧闹，这将影响到对学校态度的调查。

在考虑使用系统抽样时，首先仔细检查可能的被调查者名单。如果你怀疑抽样单位（人、教室）的排列或顺序可能带来偏倚，就使用其他抽样方法。

方便抽样

采用方便抽样是因为愿意参与调查的人在你需要他们的时候也能找到他们。比如你想了解学生健康服务如何。你计划访谈50名学生。如果你白天站在诊所门口，你可以招募每个经过的人。当你找够50名访谈对象，就有了一个方便抽样。这一抽样会在以下几个方面产生偏倚：

- 愿意接受访谈的学生可能比拒绝的学生更关注健康服务。
- 在你访谈的时间使用服务的学生可能是为了方便。而患病的学生可能晚间使用服务。
- 和你交谈的学生可能有牢骚要去抱怨。
- 和你交谈的学生可能是最满意的因而想去夸赞。

• 学生可能想和你交谈但当时没时间。他们可能是刻苦的学生。或许刻苦的学生(较为年长?)在需求和观点上不同于其他学生。

做出决定

因为存在偏倚,除非你提供其他证明,否则方便抽样不可信。以下措施可以增进该方便抽样的可信度:

• 询问拒绝者和参加者对健康的关注程度并比较回答。你可能发现没有差别。如果是这样的话,那么你的方便抽样的发现就获得了支持。

• 晚间访问诊所了解使用服务的学生的健康状况是否与白天的学生不同。你可能发现没有差别。同样,如果这样的话,你的抽样获得支持。

• 询问学生他们是否有所抱怨。

• 询问学生的满意度。

• 询问学生为什么拒绝参与。是否因为现在没有时间?

• 从年龄、性别等方面比较参与和未参与的学生。如果没有发现差别,你的方便抽样的发现相对更容易获得支持。如果发现差别,你需要重新考虑你的抽样方法。

其他方便抽样方法

考虑这些常见情形。

■ 示例　　**其他方便抽样方法**

1. 对 100 名法学院院长、大型律师事务所高级合伙人和法官进行调查,了解他们去找哪个律师解决他们自己的法律问题。结果以“世界最佳律师”为题发表在《全球新闻和世界报告》上。
2. 是什么导致大学生自由或保守? 家庭背景? 出生地? 当前的宗教信仰? 教育成就? 收入? 开展了一项针对青年保守协会和青年自由学会成员的调查。调查结果从学生的视角揭示了原因。
3. 一项流行的电影下载功能大幅度提高了某项服务的月花费。增加的价格

将会怎样影响订阅者？大多数人是会保留服务还是另寻他处？总共联系了 80 000 人完成一项网上调查。所有潜在的被调查者之前都参加过 SurveyPanels（一家互联网调查公司）开展的调查。■

在第一个例子中，"顶级"的律师可以提供院长、高级律师事务所合伙人和法官所需的服务。顶级律师对一般的法律问题的服务的合用性如何？在第二个例子中，只有两个组织的学生被调查。我们是否能够相信他们代表了所有学生，包括哪些属于其他群体或未参加任何群体的学生？在第三个例子中，调查发送给 80 000 名曾经参加过 SurveyPanels 举办的其他网络调查的人。所有的被调查者都愿意完成一个网上调查。他们是否代表了这个电影下载服务的一般使用者？

你使用方便样本是因为他们能够找到并接触，因而避免了制订并实施抽样方案所需的时间和专业知识。如果调查者确信来自样本的回答能够代表目标总体，方便样本就是有用的。如果参与者在某些重要的方面（如年龄、性别、网络知识）与目标总体显著不同就会有问题。

这些是标准的方便抽样技术：

滚雪球抽样。之前选出的群体成员推荐其他成员。例如，选择 CEO 然后请他提名其他人。

配额抽样。以特定比例将群体划分为子群。这类似于分层抽样。

焦点小组。10~20 人被聚在一起回答特定问题。一个训练有素的组长主持会议。通常有记录员在现场记录和总结讨论。

固定样本组（panels）。人们被招募起来根据需要参加调查。许多网络调查公司从参与他们调查的数百万人中招募被调查者的固定样本组。大型的有信誉的公司对被调查者进行筛选，组建起各种群组，比如说，熟练智能手机使用者、在职妈妈、视频游戏者或随机选择的美国成年人。通常，调查者设计问卷并规定谁应该被纳入调查。调查公司做其余的事情。根据回答付费。例如，如果调查者需要 300 个回答，调查公司按每个回答 3 美元收费，那么调查的总费用就是 900 美元。通常需要花费 1 周时间得到完整的回答。为了鼓励回答，调查公司通常对固定样本组成员提供奖励。例如，一家著名的网络公司根据每个固定样本组成员的意愿捐献 0.5 美元给慈善机构，并给予他或她

赢取 1 美元礼品卡的机会。公司不会直接提供款项或积分,以防止鼓励那些只对奖金感兴趣的参与者。

寻找样本:谁进?谁出?

如何寻找一个样本?在开始寻找之前,必须决定谁应该被纳入(或移出)调查。假设你希望评估 COMPULEARN(这是一个让员工掌握最新产品的电子营销方法的网络项目)的有效性。你决定调查一个用户样本,为应该纳入调查的被调查者设定了以下标准:

- 参加项目一个月或以上
- 愿意学习使用 FREENet Software(大约 30 分钟)
- 30 岁或以上
- 能读英语

这些合格标准帮助你缩小样本范围。但是它们还起到其他作用:使你的发现只适用于满足这些标准的其他人。原因是,在你的调查完成后,你只能就参加项目一个月或以上、能够使用特定软件、30 岁或以上以及说英语的人阐述 COMPULEARN 项目的有效性。一个 29 岁的人且满足其他所有标准的人是不合格的,因此你不能够把你的调查发现用于 29 岁的人。(毕竟,你选择 30 岁作为合格标准,这是一个严格的分界点。)

每次做调查,都必须决定:我希望抽取所有人还是只抽取符合某种特征的人或地方?通常的回答是"只抽取符合某种特征的人或地方(如拥有 350 名或更多学生的学校)。"

样本量多大为宜?

有些调查只涉及一组。抽选选民样本的民意测验就是这种类型的调查抽样。诀窍在于选择一个你感兴趣的所有选民的"代表"。其他调查涉及两组或更多组。例如,一项调查比较两组学生的职业规划,一组学生参与 JOBS 项目,另一组学生参与 CAREER 项目,这就是第二种抽样。当比较学生的职业规划时,你可能希望获得两个群体的

学生的代表性样本。除此之外,你还需要考虑两组学生每一组的人数,以便两者之间存在差异时你有足够的“统计效力”检出它们。

考虑这些例子:

1. **一组,无干预或项目**:你为公园系统工作,希望发现年轻人在哪里骑山地自行车。

   目标:调查青年山地自行车爱好者的偏好

   总体:21 岁或以下的山地自行车爱好者

   抽样问题:需要调查多少名山地自行车爱好者才能比较好地代表 21 岁或以下的山地自行车爱好者?

2. **二或更多组以及一项干预**:你有两组。在其中一组参与一项新活动后,你希望比较两组之间的差异。

   目标:评估一个国际象棋项目

   总体:参与一个实验性国际象棋项目的孩子和一个没有参与该项目的对照组。

   抽样问题:每组应该有多少孩子,才能在差异确实存在的情况下检出该差异?

当你考虑样本量时,你也必须考虑标准误差,这是一个描述抽样误差的统计量。误差的存在是因为当你抽样时,你从一个大的总体中抽取,而样本通常不等同于总体。这种差异是抽样的随机结果。你可以控制它,但不可能完全消除它。如果你做无限次的抽样,均值或平均数将形成一个聚集于真实总体均值周围的分布。这种分布,形状像钟形,就是所谓的正态分布。通常,较大的样本更可能靠近真实的总体均值,因而是对总体均值更准确的估计。

标准误差

当样本是随机抽取时,较大的样本能够减少抽样误差。描述抽样误差的统计量称作均值的标准误差。它是抽样均值估计的分布的标准差,该均值是进行无数次给定样本量的抽样后形成的。

尝试这个:在一个随机抽取 300 名被调查者的调查中,70% 的人对下面的问题作出肯定回答:“山地骑行是你最喜欢的运动吗?”调查者报告抽样误差是 9.2 个百分点。如果你在 70%之上加上或减去9.2,

就得到一个从60.8%到79.2%的区间——称作置信区间。使用标准的统计方法,这意味着调查者有95%(95%是个惯例)的把握估计"真实的"回答喜欢山地骑行的人的比例落在这个区间内。

要点是使置信区间尽可能小。实际上,在随机抽样中较大的样本通常会减少抽样误差。当样本较小时,增加样本而能减少的误差远多于当样本较大时。另外,不同的抽样方法(如系统抽样)可能产生不同于随机抽样的误差率。

记住,不是所有的误差都来自抽样。虽然你希望样本尽可能大,以保持误差的低水平,但你不会希望因为样本量太大而导致其他来源的误差增加。其他误差来源包括模糊的合格入选标准、设计和执行糟糕的调查和不佳的回收率。

统计方法:两组及一个干预的抽样

假设你希望比较两个组。首先,把总体一分为二(比如说,随机地)。然后,使用统计计算(如下面这种)找出每组的样本量是否大到足够让你检出差异,如果确实存在差异的话。如果总体巨大,你可能希望抽取一个样本(比如说,随机地),然后把人分派到两个组。如果你是随机抽取的样本,那么就是随机抽样。如果你随机把人们分派到组中,就是随机指派。例如你抽取一个城市里的所有5所学校,然后随机指派每所学校的所有学生到组,就是非随机整群抽样和随机指派。然而,如果你随机选择一个城市里的5所学校,指派3所到实验组,指派2所到对照组,然后把实验学校的所有学生纳入实验,那么就是随机整群抽样加非随机指派。

当你有两组和一个干预时,使用类似下面示例的清单获得或评估样本量。

■ 示例 **抽样两组和一个干预的样本量计算**

集中和分类调查目标和问题

确定调查的目的。考虑这些:

调查1:生活质量

目标:确定年长或年轻的妇女在乳腺癌手术后的生活质量是否有差异

问题:乳腺癌手术后年长或年轻的妇女的生活质量是否有差异?

调查2:在学校里的焦虑

目标:确定与学校相关的焦虑的性质和类型

问题:男孩女孩的焦虑是否有不同?年轻与年长的学生相比如何?

■

每个目标或问题都包含自变量和因变量。自变量被用来预测或解释因变量。自变量通常包括被调查者所属的组(实验组或对照组,男性或女性)或他们的特征(50岁以下和51岁及以上)。以这个问题为例:"男孩和女孩的焦虑是否有所不同?",分组或自变量是性别。

因变量是调查测量的态度、属性、行为和知识。用统计学术语来说,它们是有待估计或做出推论的变量。在这个问题中,"男孩和女孩的焦虑是否有所不同?"因变量是焦虑。

识别子组。子组指的是获得其调查结果必须有足够数量的样本以确保结论准确。在前面两个调查中,子组的识别可以通过查看自变量。调查1的子组是年长和年轻的妇女。调查2的子组是年长和年轻的男孩和女孩。

识别调查类型和数据收集需求。因变量告诉你了调查的内容。例如,调查1的具体问题将提问被调查者关于他们的生活质量的各个方面。调查2将提问焦虑。假设调查1是面对面访谈,而调查2是自填问卷。

检查调查的资源和日程。一个有许多子组和测量的调查将比很少子组的调查更昂贵。考虑这个:

[子组、测量、资源和日程]

调　查	子　组	调查类型	内　容
1.乳腺癌手术后年长或年轻的妇女的生活质量是否有差异?	年轻和年长妇女:两个子组	面对面访谈	可能需要时间去雇用和培训针对年轻和年长妇女的不同的访谈人

续表

调　查	子　组	调查类型	内　容
			可能难以找到足够数量的符合条件的年轻或年长妇女
2.男孩女孩的焦虑是否有不同?	男孩和女孩、年轻和年长:四个子组	自填问卷	可能需要时间把问卷从英语翻译到其他语言
年轻与年长的学生相比如何?			必须采取措施确保回答的保密性

子组的数量从 2 到 4。执行、计分和解释一个组的调查就已经很困难了,多于一个组时难度还会上升。

计算样本量。假设一个调查关注一个弹性工作项目是否会增进员工的满意度。再假设调查的一个目标是比较项目内外员工的目标和抱负。每组员工的数量以多少为宜?为了回答这个问题,必须回答另外 5 个问题。

确定样本量时必须回答的 5 个问题

1.什么是零假设?

零假设(H_0)是一个关于两组的均值或平均数不存在差异的陈述。例如,员工满意度调查的一个零假设是这样的:

H_0:参与项目和未参与项目的员工的目标和满意度之间不存在差异。

2.与零假设相关的涉及总体均值(μ_o)的显著性的期望水平(α 水平)是多少?

检验开始前选定的显著性水平被称作 alpha 值(以希腊字母 alpha:α 代表)。alpha 给出了当零假设为真时拒绝之的概率。一般希望 alpha 尽可能小——0.05、0.01 或者 0.001——以避免拒绝为真(组间均值不存在差异)的零假设。*p* 值代表观察到的结果(或统计检

验的结果)仅由于偶然(而不是参与项目)而出现的概率。它是统计检验后计算出来的。如果 p 值小于 alpha,那么零假设就被拒绝。

当研究发现两组之间存在差异,但实际上并没有差异,就被称为 alpha 或 I 类错误。当没有发现组间存在差异,而实际上却存在差异,这种错误被称为 beta 或 II 类错误。

3.有多大机会检测到真实的差异?

效力是指如果差异真实存在而检出给定大小的差异的能力。它的计算是 $1-\beta$(希腊字母 beta)。它被界定为如果零假设错误时拒绝零假设或者备择假设为真时接受备择假设的可能性。人们总是希望有高的效力。

4.多大的均值差才是重要的? 也就是说,有意义的 $\mu_1-\mu_2$ 是多大?

假设调查使用的是目标及满意度量表(GASS)。这个假想的量表有 50 分。如果量表是有效的,你将可以获得公开的计分规则和描述分数意义的数据。提出下列问题:是否分数越高越好? 多少“点”产生有实际意义的差异?

回答此类问题将帮助你确定希望在两组间检出多大的差异。

5.什么是对总体的标准差 σ 的良好估计?

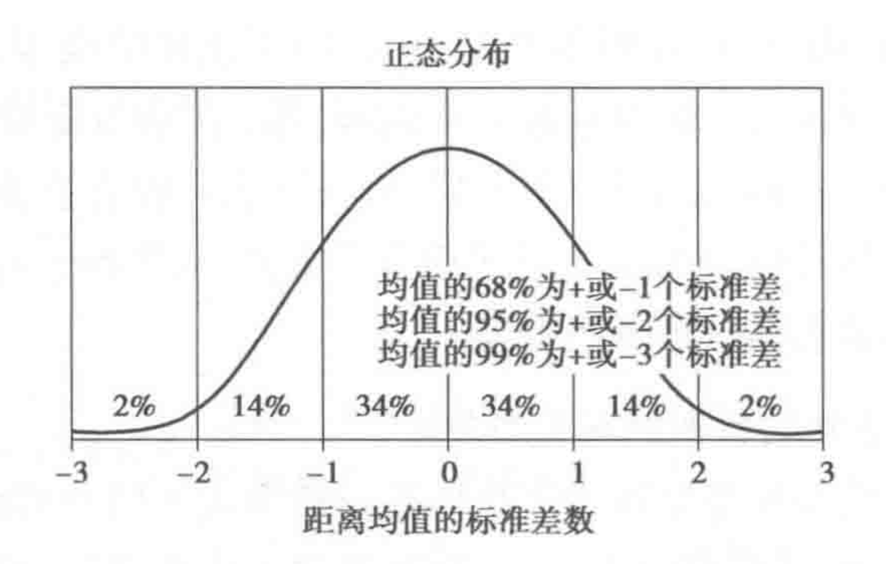
标准差(σ,小写的希腊字母 sigma)是对均值数据的散布或极差的一种常用测量。

如果观察或数值的分布是钟形或正态分布的,那么 68%的观察将落在均值±1 个标准差的范围里,95%的观察落在±2 个标准差的范围里,99%的观察落在±3 个标准差的范围里。看下图:

标准差的估计可以来自前次完成的调查。检查用来计算标准差的样本是否与你自己的样本类似。如果答案为否,则你组中的标准差可能是不同的,组也是不同的。

另外一种估计你组中标准差的方式是开展一个 25~50 人的小型试调查。你还可请专家给你最高和最低值的估计,以此为基础计算标准差。

接下来是一个计算样本量以比较两个独立组(如两个员工组)均



值的公式。组1参与了一个提升满意度的项目,而组2未参加。这个公式假设两组的样本量和标准差是相同的:

$$\frac{(z_{\alpha} - z_{\beta})\sigma^2}{\mu_1 - \mu_2}$$

式中　$\mu_1 - \mu_2$ 是两组间检出的差异的大小。

z_{α}(正态分布中的上限)和 z_{β}(下限)被界定为:

$$z_{\alpha}\frac{X - \mu_1}{\sigma/\sqrt{n}} \text{ 和 } z_{\beta}\frac{X - \mu_2}{\sigma/\sqrt{n}}$$

以下是一个如何应用公式的例子。

■ 示例　计算一个员工调查的实验组和控制组的样本量

调查情境。Devsoft 的员工正在参加一个提升其工作满意度的项目。在3年期项目结束时,将对实验组和控制组的参与者进行调查,以了解他们的目标和抱负。调查中最高的可能分数是100分。alpha水平设定为0.05。检出真实差异的可能性设定在0.80。一组满意度测量专家认为实验组和对照组的满意度测量分差应该在10分或以上。之前的员工调查发现标准差为15分。

计算。为了计算,假设分布是标准正态分布或 z 分布。标准正态曲线的均值为0,标准差为1。(更多关于标准正态分布、单双尾检验和 z 值的内容参见第6章。实际的 z 值可以从统计教科书或程序手册中获得。)与 $\alpha = 0.05$对应的双尾 z 值为+1.96。对于 $\alpha = 0.01$,双尾 z 值为2.58。对于 $\alpha = 0.10$,z 值为1.65。对于 $\alpha = 0.20$,z 值为1.28。与 β 相关的单尾 z 值的下限是-0.84(临界值或 z 分数将 z 分布的最低20%与80%分隔开)。应用这个公

式,我们有如下计算:

$$(1.96+0.84)(15)^2 = 2\left(\frac{42}{10}\right)^2 = 2(17.64) \approx 36$$

每组至少应该有 36 名员工才能有 80%的机会检出 10 分的差异。

有时候你可汇集某一特定数量的人参加调查。你怎么才能知道这个数量是否大到能够发现差异?这种时候,统计方法可以帮到你。看下面。

■ 示例 **检出差异的效力**

戒酒中心有一个减少酗酒风险的项目。它设定的成功标准是:

在项目结束时,在治疗组中 20%的有害饮酒者将减少风险,而在对照组中这一比例为 10%。

关于他们的风险,中心希望每组调查 150 人。这一样本是否足够大?一位统计学家被邀请回答这一问题。统计学家给出下表,使用 $\alpha=0.05$ 的双边检验比较百分比的变化。

N 或样本量	%1	%2	效　力
50	20	10	0.29
100	20	10	0.52
200	20	10	0.81
250	20	10	0.89
300	20	10	0.93

这张表可做如下解释:

如果实验组 150 人(最终分析的样本量)对照组 150 人,我们将有 69%的效力分辨实验组中 20%的风险减少与对照组中 10%的减少。 ■

你可以使用统计方法区分不同的效应量。效应量是差异除以标准差。看下面。

■ 示例 **80%的效力和效应**

戒酒中心有完整的调查数据,其中实验组 150 人,对照组 150 人。这些人将在参与中心项目前和刚刚结束时接受调查。这个样本量能产生多大的

效力?效力是检出差异或效应的能力。换句话说,它是一种统计检验,当备择假设为真(零假设被拒绝)时检出特定大小差异(组间)的备择假设的能力。统计学家给出下表。该表意味着什么?

N 或样本量	效应量
50	0.56
100	0.40
150	0.32
200	0.28
250	0.25
300	0.23

可以对该表做如下解释:

如果实验组有 150 人(最终分析的样本量)对照组有 150 人,我们能够区分实验组的历时性变化与对照组的历时性变化之间 0.32 的效应量。

■

因为你是通过将均值或平均数的差除以标准差来估计效应的,你可以看下,举个例子来说,实验组和对照组均值增加的标准差是否分别为 0.50 和 0.10。如果是的话,效应量就是 50-40=10>0.32。在这个例子中,你至少有 80%的效力去检测实验组和对照组之间 0.40 的差异。

前面的讨论帮助你了解相关术语,并帮助你评估你的抽样方法和结果的有用性。样本量、效应和效力是统计术语。除非你准备去学习统计学,否则就请个专家帮你。或者,考虑使用统计软件或网上提供的样本量计算器。访问你最熟悉的搜索引擎键入"样本量",就能够找到免费的样本量计算器。

以下说明如何确定和验证样本量、效应和效力。

■ 示例　　样本量、效应和效力

1.瑜伽与慢性背痛

英国 BEAM(背痛训练和治疗)组织的试验发现在 Roland-Morris 功能障

碍问卷得分上 1.57 分的变化是一个比较理想的差异。假设标准差是 4 分(对英国 BEAM 组织的样本量而言),那么效应量就是 0.39(1.57/4)。为了检测这一效应量,假定 80%的效力和 20%损耗下,需要总量 262 名被试(每组 113)。

如果你想更多了解这一研究,参见 Tilbrook, H. E., Cox, H., Hewitt, C. E., Kang´ombe, A. R., Chuang, L.-H., Jayakody, S. et al.(2011). Yoga for chronic low back pain: a randomized trial. *Ann Intern Med*, *155*(9), 569-578.

2.电话辅导对父母戒烟的有效性

基于类似的研究,我们期望在 12 个月的评估期内,电话辅导组和对照组的七日戒烟率的差异为 6%(分别为 13%和 7%)。假定统计功效的目标值是 0.80。以双边 0.05 的显著性水平对假设进行检验。根据中途可能流失的被试校正计算得出样本量。除此而外,还对样本量进行调整以便能灵活机动地做出额外分析。

如果你想更多了解这一研究,参见 Schuck, K., Otten, R., Kleinjan, M., Bricker, J., and Engels, R.(2011). Effectiveness of proactive telephone counselling for smoking cessation in parents: Study protocol of a randomized controlled trial. *BMC Public Health*, *11*(1), 732. ■

回答率

回答率是回答人数(分子)除以应回答人数(分母)。如果应回答人数为 100 人,有 75 人完成调查可供分析,回答率就是 75%。

所有调查都期望一个高的回答率。然而,没有哪个回答率被认为是标准的。在有些调查中,95%和 100%是期望值;而在另外一些调查中,70%就够了。

以下是提升回答率的一些忠告。

[提升回答率的忠告]

- 了解你的被调查者。确保他们能够理解问题,重要的是,不要轻慢他们的社会文化价值。
- 使用受过训练的人招募被调查者和开展调查。建立一套质

量保证体系以监控质量和再培训。

- 多多准备超出你所需的被调查者,以防没能达到期望的样本量。注意防止超支。
- 只有当你确信被调查者对主题感兴趣时再进行调查。
- 为调查回答保密或匿名。
- 向未回答者寄送提醒函。
- 为完成调查提供可选项。例如,有的人喜欢网上调查,而其他人喜欢纸笔调查,通过邮件或传真完成和返回调查。
- 提供礼品或现金奖励。
- 制定现实的合格标准。
- 事先估计因为调查环境(如不正确的地址)或偶然因素(突然生病)而不能参与调查的被调查者的比例。
- 告诉被调查者你将如何使用调查数据。
- 准确描述如何保护隐私。

加　权

调查很少从感兴趣的总体中的每个人那里获取信息。为了避免不一致并确保调查结果能够代表总体,调查者经常为每个个案赋予权重或数值。假设你调查了100位年龄在15~64岁的人,其中60女40男。对于这个年龄段,你应该从50男50女那里获得回答。为了校正这一差异,你可以为每个男性的回答赋予一个权重,使其回答比单个女性的回答贡献更大(权重更大)。

两种最常用的加权方法是设计权重和后分层或无回答权重。设计权重用来补偿特定个案的"过抽样"或"欠抽样"或不合比例的分层。比如说你打算比较A组的50人和B组的50人。你知道A组有许多经常拒绝接受调查的人。你决定从A组中多抽(过抽样)一些人,希望即使拒绝率相对较高,最后也能使两组接受调查的人数大致相同。如果A组的抽样比例比B组高4倍,那么A组的设计权重就是1/4,B组为1。

后分层或无回答权重用来补偿具有某种特征的人没有机会作答。大多数针对一般人群的调查,女性被调查者会多于男性被调查者,虽然总体中通常是男性较多。因为调查会"过代表"女性而"欠代表"男性,所以你要赋权以弥合这种偏倚。有许多被调查者特征可能导致回答率高,包括年龄、教育、种族/民族、性别和居住地。使用后分层权重的难度通常比使用设计权重的难度大,因为需要关于总体的更多信息。这意味着需要查询和分析非常大的数据库(如国家的人口普查或大学的录取名录)以获得你感兴趣的人口变量的估计值。这些数据库提供了诸如以下这些问题的答案:总体中真实的男性和女性比例是多少?这种病的流行程度在年轻人和年长者中有何不同?生活在单亲家庭中的人的比例与其他家庭形式中的比如何?

这里是一个计算后分层权重的假想的例子。

■ 示例

性别	总体中的比例*	样本中的比例**	总体/样本	权重***
女	0.5	0.6	0.5/0.6	0.833
男	0.5	0.4	0.5/0.4	1.25
合计	1	1		

* 人口普查提供的总体中的男性和女性比例信息。

** 调查的被访者当中有60%的女性。

*** 得到的比例与总体的比例不同,女性被调查者的比例更高。赋予男性一个较大的权重以弥合差异。

■

小　结

三种最常用的随机抽样方法是:

1.简单随机抽样

2.分层随机抽样

3.简单随机整群抽样

随机抽样给调查中每个合格的参与者以公平的被选中的机会。

你可以抽取个人或更大的单位,诸如学校、办公室和医院。这些

较大的抽样单位包含群集(学生的、员工的、护士的、医生的、病人的),所以被称作整群抽样。

非随机抽样比随机抽样更容易找到被调查者。但是这种便利可能是以牺牲结论的普适性为代价的。

对于系统抽样,从单位名单中每 n 个单位抽取 1 个。如果 n 是随机选择的,系统抽样类似于随机抽样。

在方便抽样中,你选择你需要他们时能找到并且符合你的调查标准(合适的年龄或阅读水平、在最近一次选举中投过票、在社区至少生活 1 年,等等。)的每个人。其他非概率抽样方法包括滚雪球和配额抽样。

其他非随机抽样策略包括:

滚雪球抽样。之前找到的组成员推荐其他成员。例如,你抽选 CEO 然后让他们提名其他人。

配额抽样。将组以特定比例划分成子组。这与分层抽样类似。

焦点小组。召集 10~20 个人回答特定问题。一个有经验的组长主持会议。通常有 1 名记录员在现场记录讨论。

固定样本组。被调查者在需要的时候参与调查。被调查者是按照他们的人口和其他特征(如,经常网上购物、从未用过 Twitter)事先筛选出来的。

当样本是随机选取的,大样本能减少抽样误差。描述抽样误差的统计量被称为均值的标准误差。它是均值的抽样估计的分布(经过无数次给定大小的抽样形成的)的标准差。

调查很少从目标总体中的每个人那里获取信息。为了弥补损失,确保调查结果能够代表总体,调查者为每个个案赋予权重或数值。有两种类型:设计和后分层或无回答权重。

思考一下

1.找出后面文献中的调查然后做如下事情:

a)讨论调查者使用的抽样类型。

b)通过列出被纳入和被排除的被调查者的特征,描述调查者的调查对象入选标准。

2.任意找 3 个属于大型研究一部分的调查,讨论其抽样方案。抽样方案的讨论是否足够详细?也就是说,你是否有足够的关于抽样策略和样本量信息以至于在必要的时候可以重做调查?你是否还需要其他什么信息?

3.列出至少 5 个提升回答率的忠告。

4.解释这些术语在谈论调查时是如何被使用的:标准误差、置信区间、效力、效应量、显著性水平和正态分布。

文 献

Holman, E.A., Silver, R.C., Poulin, M., Andersen, J., Gil-Rivas, V., & McIntosh, D.N. (2008). Terrorism, acute stress, and cardiovascular health: A 3-year national study following the September 11th attacks. *Archives of General Psychiatry*, 65 (1), 73-80.

O'Connor, T.M., Yang, S.-J., & Nicklas, T.A. (2006). Beverage intake among preschool children and its effect on weight status. *Pediatrics*, *118* (4), 2005-2348.

Schuck, K., Otten, R., Kleinjan, M., Bricker, J., & Engels, R. (2011). Effectiveness of proactive telephone counselling for smoking cessation in parents: Study protocol of a randomized controlled trial. *BMC Public Health*, *11* (1), 732.

Skoffer, B. (2007). Low back pain in 15-to 16-year-old children in relation to school furniture and carrying of the school bag. *Spine*, *32* (24), E713-E717.

Tilbrook, H.E., Cox, H., Hewitt, C.E., Kang'ombe, A.R., Chuang, L.-H., Jayakody, S., et al. (2011). Yoga for chronic low back pain. *Annals of Internal Medicine*, *155* (9), 569-578.

Wehr-Flowers, E. (2006). Differences between male and female students' confidence, anxiety, and attitude toward learning jazz improvisation. *Journal of Research in Music Education*, *54* (4), 337-349.

5

调查设计:环境控制



可以对一组或多组人一次或多次进行调查。例如,你可以在一所学校的学前儿童参加一个新的音乐项目之前和刚刚结束时进行调查,或者对10所学校的学前儿童在参与一个项目之前进行调查,并在其后每2年调查一次,直至他们中学毕业。调查的设计指的是其频度(一次或更多次)、次序(一个新项目刚开始或刚结束时)以及涉及的组数(一个学区的所有学生或者10个学区的所有学生)。

截面设计提供事物在某一时点的画面。某项选举前1个月对投票人偏好的调查,以及在某个调查中投票的人的年龄和收入的调查,都属于截面设计。

纵向调查被用来寻找变化。如果你在2012年抽取一个6年级孩子的样本,然后在接下来的5年里每年跟踪调查,你使用的纵向设计被称作队列设计。

其他研究设计包含比较。如果被调查者被指派到一个随机形成的组,这就是一个真正的实验。如果你使用随机数表或统计算法指派250名工人到一个实验项目,指派250人到对照组,并调查他们的满意度,就具有了真正的实验的潜质。真正实验还指随机对照试验。这些设计能够验证项目

A是否导致结果A,特别是当调查组是随机选择和指派的时候。理想状态是“盲化”所有的参与者,让他们不知道是在实验组还是对照组,不过说得容易做到难。

准实验设计是一种由志愿者或根据方便抽样形成的比较组设计。假设10个班级同意参加一个关于创新性阅读方法的研究。如果你将这10个班级与另外10个仍然参加常规阅读项目的班级相比,就是一种准实验设计。对于这类设计,调查者必须警惕一开始就存在的动机和教育的组间差异。

个案控制设计是根据具有还是不具有被研究的状况进行分组(个案组和控制组),然后就被认为与引发状况的原因相关的现在或过去的态度、习惯、信念或人口因素进行组间比较。例如,当你希望比较一周或更长时间读1本书的人和不读书的人,他们在受教育方面是否有所不同?是否孩童时代在家中接触到书?多大年龄?这就是个案控制设计。

有效的调查结果来自有效的调查(测量有效性)和有效的研究(内部和外部效度)。有效的设计取决于下列因素:选择调查参与者的客观性、他们的回答随着时间推移的稳定性、被调查者的心理和身体成长以及外部世界的事件。如果不了解也不解决这些因素,就会影响调查设计的内部效度。当这些因素影响到把研究发现应用于其他场景和其他被调查者的效果时,调查结果的外部效度或概括性就受到威胁。

可以采用哪种设计?

调查可以单独进行。一家公司对员工满意度的民意测验是单独调查的一个例子。但是调查也经常被作为一个大型研究的一部分,该研究可能还有其他数据收集方法。假设一个公司对提高员工绩效感兴趣。公司设立了一些新项目,然后将部分员工随机分配进去,而其他人仍旧照常。6个月后,公司访谈两组员工以了解他们对两个项目的满意度,观察他们的工作表现,评估其工作期间的绩效。这其中只

有一种测量方式——访谈,真正属于调查。但不管怎样,调查者需要确信评估新项目的研究设计符合调查的需求。

调查数据可以用来描述事物的状况、显示变化和进行比较。调查设计指的是控制或组织“环境”的方式。环境指的是被调查者的数量和特征以及多久和何时进行调查。对环境的控制越多,你的结果也越准确。

调查者控制最多的环境变量如下:(a)何时进行调查(如刚毕业时),(b)多久(如,两次:刚毕业和毕业 12 个月后),(c)样本量(如 10 名或 1 000 名毕业生),(d)组的数量(如,只有 1 组,所有的毕业生;或不止一组:公立学校毕业生和特许学校的毕业生)。

看一看由 Have-A-Heart 协会(以下简称 HAH 协会)规划的这些调查。

■ 示例 **不同设计的调查**

1.HAH 协会向社区居民提供 10 个教育项目。在 6 月份,它对一个随机样本开展一项调查,了解和描述人们选择了哪个项目以及原因。

2.HAH 协会希望了解参与其教育项目的人们是否获得了有关减轻心脏病风险方面的知识。协会对 2009 级的 300 名志愿参加者进行了一次调查。这项调查每 3 年举行一次,对每次的结果进行比较以看一看知识是否能够保持。

3.HAH 协会对两个旨在通过教育减少心脏病风险的竞争性项目的有效性进行了评估。参与者是随机从协会的数据库中选出的。他们被随机指派到两个项目之一。在参加项目前和之后的 6 个月和 12 个月,调查了参加者的心脏病风险。 ■

第一个 HAH 协会调查于 6 月举行旨在描述社区选择的项目。其设计,在一个时间点上收集数据,被称为截面设计。

第二个 HAH 协会调查为了收集一个班级 3 年期间的信息。这是一种纵向、时间序列或队列设计。

第三个 HAH 协会调查作为评估两个项目有效性的一部分而使用。它使用的是对照组设计。

表 5.1 显示了 HAH 协会的三个调查的目的、抽样和设计要点、结

果和设计类型之间的关系。

表 5.1 HAH 协会的三个调查的目的、抽样和设计要点、结果和设计类型之间的关系

调查希望了解什么	抽样要点	设计要点	结果	设计类型
对教育项目的偏好	完成项目者的一个随机样本	何时开展:今年	项目选择	截面
从教育项目中获得的知识	来自 2009 级的 300 名志愿者	何时开展:2009、2010、2011	知识	纵向:队列
项目的有效性	从数据库中随机选择然后随机指派到项目	何时开展:3 次,开始参与以及其后的 6 和 12 个月 多少组:2 组	心脏病风险	对照:真正实验(随机对照试验)

截面调查设计

对于这种设计,数据是在一个时点上收集的。可以把截面调查想象成对人群或组织的一张快照。

假设 HAH 协会想了解社区偏好它的哪个教育项目。考虑这个问题及其回答:

■ 示例

截面设计

问题:如果你可能有一个项目,你选择哪个?

样本:从当年参加过教育项目的人中随机抽取的 500 人的截面调查

设计:截面

方法:电话访谈

回答: Dine Out(外出就餐)项目胜出。结果见表 5.2 和表 5.3。

表 5.2 男性和女性参与者最喜欢的教育项目

	男		女		合 计	
	人数	百分比	人数	百分比	人数	百分比
“外出就餐”项目	168	34	175	35	357	69
“心情舒畅”项目	75	15	50	10	97	25
“紧急护理”项目	21	4	11	2	46	6
合 计	264	53	236	47	500	100

表 5.3 不同年龄的参与者最喜欢的教育项目

	21~45		45~65		66+		合 计	
	人数	百分比	人数	百分比	人数	百分比	人数	百分比
“外出就餐”项目	99	20	96	19	41	8	236	47
“心情舒畅”项目	35	7	37	7	90	18	162	32
“紧急护理”项目	47	9	52	10	3	1	102	21
合 计	181	36	185	36	134	27	500	100

假设已经根据某种随机抽样技术抽选了一个参加者样本并且已经正确提问,本例中的表格显示“外出就餐”项目是获胜者。这是因为:

1.无论性别或年龄,“外出就餐”项目都领先。

2.更多男人比女人喜欢“心情舒畅”项目和“紧急护理”项目。但当到了“外出就餐”项目时,男人和女人的偏好差不多。

3.65 岁以上的人喜欢“心情舒畅”项目,但在另外两类中的人数没有那么多。

当然,你可能希望只使用表 5.2 或只使用表 5.3 来说明你调查的一个观点。你也可以把它们结合在一个大表里。但是一个仔细规划的截面设计给你多种方式分析和呈现你的调查数据。

对于这个项目偏好数据,例如,你可能还想考虑职业:商人与其他职业相比如何?退休人员是否有所不同?或者你还考虑居住地:城市

某区的人与其他区的人相比如何?

做出决定

截面调查有重要的用途。他们如实描述事物,因而人们可以规划改变。如果他们对截面调查揭示的画面不满意,可以改变它。截面调查也相对容易做。然而,它们又是有局限的,因为如果事物快速变迁,调查信息可能很快会过时。

纵向调查或队列

纵向调查设计指的是收集不同时间的调查数据。有时也称为时间序列设计。队列设计指的是不止一次地调查一个特定组(如六年级学生)。看下面的例子中的被调查者,他们参加了一个由 HAH 协会资助的项目。

■ 示例　　　　队列设计

问题:参加者关于心脏病了解什么?

样本:一个参与了"外出就餐"项目的 300 名志愿者的样本。该项目旨在教给人们心脏病的原因、良好营养以及在外出就餐时如何避免心脏病。

方法:网上和邮寄自填问卷。

回答:在 3 年的调查期间,参与者始终不太了解疾病预防。到 2009 和 2010 年,他们对营养有所了解,到 2011 年他们对心脏病的原因了解了很多。

证据显示在表 5.4 中。

表 5.4　参与者的心脏病知识(N=1 500)

年　份	病　因	营　养	预　防
2009	很少	很少	很少
2010	有些	有些	很少
2011	很多	有些	很少

来源:心脏病信息量表的分数,健康调查基金会,© 洛杉矶,2012 ■

做出决定

队列为调查者提供了效应(比如说,像“外出就餐”这样的项目)是否持续的信息。队列设计的问题是,除非被调查者受到鼓励,否则可能会退出,剩下的人可能不能代表整个组。另外,群组可能是不典型的,因为组成的人是愿意并且有时间参与项目并完成调查的。

对照组调查设计:准实验和真实验

对于这类设计,人们被分为两组或更多组,然后对他们的调查结果进行比较。组间的差异必须在他们被调查前就已经知道,或者其中一组必须参与一个项目因而有所不同。典型的对照组设计将“实验”组与“安慰剂”或“常规对待”组相对比。考虑下面的项目:

1.开展一项关于投票人对校董事会候选人的偏好的调查。作为分析的一部分,对男性和女性的偏好进行了比较。

2.锻炼的频率和类型与头疼的严重程度之间存在怎样的关系?区分了两个组。第一组经常头疼,而第二组很少或从不头疼。对两组进行调查,了解他们的锻炼习惯,然后比较结果。

3.如何让人们参加校董事会选举?选举前1个月,招募了两组志愿者。第一组播放一段视频录像,第二组由社区名人演讲。选举后,对两组进行调查比较投票的人数,并让其解释他们的原因。

项目1和2使用了对照组设计。项目2中的两个组是事先已知的,而项目3中的组则是为了比较的目的而专门设立的。项目1使用的设计是一种截面设计。对男性和女性在一个时点上的偏好进行了比较。

对照组设计有时分为准实验和真实验设计。在准实验设计中,指派到组通常是故意的和非随机的。在真实验(有时称为随机对照试验)设计中,个体被指派到组是偶然的结果。真实验是两者中更有效力的,因为如果正确实施的话,你可以有把握地说项目A引起了结果A。

■ 示例 **一个准实验设计**

你被要求评估两个老年人项目。符合条件的参加者分配到项目中采取先来先服务的原则,结果项目 1 有 60 人,项目 2 有 59 人。调查解决的问题之一是参加者对其项目的职员是否满意。为了回答这一问题,你要求两组人在项目结束后 3 个月完成一个问卷。这个调查的设计类似于:

两个老年项目参与者对职员的满意度

项目 1(N=60)	项目 2(N=59)
平均态度分在此	平均态度分在此

来源:数据来自对职员的满意度量表的分数。 ■

用来调查两个项目中的老人对职员态度的准实验对照组设计的有效性如何?考虑这些可能性:

1.两组的参与者可能在项目开始时就是不同的。例如,老年人可能更喜欢其中一个项目。

2.某个项目中真正不喜欢职员的参与者可能会退出项目。

■ 示例 **对照组与纵向设计**

两个老年项目提出的另外一个问题是参与者是否了解常用药。为了回答这一问题,参与者被访谈了 3 次:开始、第一个月结束和第一年结束。这种调查设计策略可以描述如下:

两个项目中常用药知识的变化

时　间	项目 1(N=60)	项目 2(N=59)
项目开始	访谈结果在此	访谈结果在此
第一个月结束	访谈结果在此	访谈结果在此
年末	访谈结果在此	访谈结果在此

来源:数据来自对每个项目参与者的访谈 ■

如果具有严重健康问题的人有更多机会被指派到某个项目,或者有更大几率退出某个组,就会威胁到这一设计的效度。

■ 示例 **真实验对照组设计**

政府开展一项调查以确定 3 个老年项目中哪个的作用最大。关注点之

一是项目的成本。采用对照组设计,将 Freda Smith 中心的人随机指派到 3 个项目之一,然后对 3 个项目的成本进行比较。项目 1 有 101 人,项目 2 有 103 人,项目 3 有 99 人。

Freda Smith 中心 3 个老年项目的成本比较

成　本	项目 1(N=101)	项目 2(N=103)	项目 3(N=99)
平均成本在此			
平均成本在此			
平均成本在此			

来源:数据来自对财务专家的访谈

注意:参与者是随机指派到组的

这种设计是相对有效力的。因为人们是被随机指派到每个项目中,任何与项目的影响共存的变化因素对于 3 个组都是相同的。然而,虽然人们是被随机指派到 Freda Smith 中心的 3 个项目之中的,但是其他中心可能有所不同,因此这个调查的发现可能不适用于其他地方。

■ 示例　**真实验对照组设计与纵向设计**

Freda Smith 中心的项目 1 和 2 的性价比被证明是一样。随后政府开展一项研究以确定哪个项目被参与者认为提供了更好的医疗保健。为了做出评定,采用了对照组设计,在项目开始和结束时对保健服务进行了评估,并且对项目 1 和 2 中的人进行了比较。下面的表描述了这一设计。

两个老年项目中参与者获得的医疗保健的比较

时　间	项目 1(N=101)	项目 2(N=103)
项目开始时		
项目结束时		

来源:数据来自 ABY 保健质量评论。体系——对医生、护士和病人的调查。

这个真实验和纵向设计是最精密的了,能够让你做出好的推论。

析因设计:特殊案例

析因设计能够让研究者评估提供一种新治疗或实践的不同方法,

以便观察哪种组合最奏效。假设你正在从事一项研究,以找出是否通过事先向可能的调查对象发送电邮通知和/或恳求他们参与调查来提升网上调查的回答率。你推荐使用 2×2 析因设计,其中某些被调查者通过电子邮件通知参与调查,而另外一些则没有。另外,有些人被恳求参与调查,而另外一些则不然。因素(它们也是因变量)是恳求(因素 1)和通知(因素 2)。每个因素有两级:恳求对未恳求和事先通知对事先未通知。

		因素 1:恳求状态	
		恳求	未恳求
因素 2:通知状态	事先通知		
	事先未通知		

在 2×2 设计中,有 4 个研究组:(1)事先通知电子邮件和恳请电子邮件,(2)事先通知电子邮件和无恳请,(3)无事先通知电子邮件和有恳请电子邮件,(4)无事先通知电子邮件和无恳请电子邮件。在上面的图表中,空格中应填入的是每一类被调查者的人数(例如,恳求×事先通知组的人数对恳求×未事先通知组的人数)。如果被调查者是随机指派到每个组的,你就会有一个随机对照试验或真实验。

通过这种设计,研究者能够研究主效应(恳求对未恳求)或交互效应(事先通知和恳求)。在这项研究中的结果总是回答率。

析因设计可以包含许多因素和等级。正是根据等级数对设计命名的。例如,在一项对门诊病人、住院病人和日间治理环境中的心理治疗与行为改变的研究中,有两种因素(治疗和环境),其中一个因素有两级(心理治疗对行为改变),一个有三级(住院病人、日间治疗和门诊病人)。这一设计就是一个 2×3 的析因设计。

其他调查设计:常模和个案控制

两种较少见的调查设计是常模和个案控制。两者都通过使用特殊的对照组来对调查环境施加某些控制。

做出决定

真实验是以随机选择和指派为特征的。许多研究者断言最好的真实验是那种不论调查者还是参与者都不知道组别。在调查的“真实”世界或社会研究中开展真实验可能是困难的。它们相对昂贵并且需要时间去执行和分析。然而,如果你希望知道一个实验是否有效,真实验或随机对照试验是正途。

准实验设计更容易执行。然而,因为组不是随机构成的,他们初始的差异可能是原因或与结果相混淆。调查者需要尽一切可能防止这种初始的差异,比如说,根据可能影响结果的因素或变量来匹配组。变量可以包括教育、动机、收入等。可以用统计方法来帮助控制这些差异。

常模调查设计

常模是比较组的标准。常模来自现有数据或类似政府登记部门维护的大型数据库。例如,把某个学校的阅读分数与一个全国教育数据库中找到的全国平均分相比较,就是使用了常模。看下面的例子。

■ 示例

常模设计

对洛杉矶青年项目的参与者进行了 EFFICACY 调查,这是一个测量自我效能的工具,已经得到了一个全国 5 000 人样本的验证。使用全国样本作为常模,因为没有理由相信洛杉矶的组与全国其他地方在自我效能上有不同。青年项目分数因而可以与从全国样本中获得分数相比较。

常模设计——与一个范例相对照

一家新学术性医疗中心的医生对工作的满意度是否与老中心的医生相似？新中心和老中心有着同样的任务。为了回答这个问题,对两个中心的医生进行了调查,新中心被与典型的老中心相比较。 ■

做出决定

常模调查设计比其他对照设计更为便宜和省时,因为调查者不需要找到或创造一个对照组。这种设计依赖合适数据库的可得性。记

住,你的组和"常模"组可能在重要的方面事实上有所不同,导致不太有效的发现。例如,假设洛杉矶青年项目的参与者比全国样本年轻。如果年龄和自我效能相关,随着成长会有更高的调查分数,那么青年项目看起来就需要更加努力才能成功。如果你使用常模设计,准备好为你的比较常模或标准辩护。

个案控制设计

个案控制设计是根据具有还是不具有被研究的状况进行分组(个案组和控制组),然后就被认为与引起状况的原因相关的现在或过去的态度、习惯、信念或人口因素进行组间比较。

个案控制设计被检验某一特定假设的研究者所广泛使用——例如,肺癌与吸烟习惯之间存在的关系。

个案控制设计需要两组:一个个案组和一个控制组。理想情况下,两组的差异只在于个案组具有所研究的特征或状况而控制组没有。

大多数时候个案控制设计意味着选择一个控制组,控制组应当与个案组在极有可能影响状况的方面一致。如果你有证据表明性别影响头疼的频率,那么对头疼者的研究应该有男女性别比例相同的个案组和控制组。

做出决定

个案控制设计的主要缺点是,无论如何选择或匹配,两个调查组都可能完全不相似,因为不可能控制影响状态的所有特征。例如,有些匹配标准可能是不正确的。以下是一个使用个案控制设计的调查。

■ 示例　　**个案控制设计**

某大学门诊部随机将 1 000 名年龄在 21 到 55 岁的病人分配到两个组:在过去 12 个月里发生过偏头疼的人和没有发生过的人。他们使用一种分层随机抽样技术从每个组中随机选择 100 人,其中半数是女性。开展了一项调查了解每组的下列特征:

- 典型的每日活动
- 潜在的压力源

- 家庭史和背景
- 医疗史
- 饮食

■

调查设计效度

有效的调查信息来自可靠且有效的调查测量工具(测量效度)以及调查发生的环境(设计效度)。这两类效度的区别在于,测量效度指的是调查测量工具的特征,而设计效度指的是调查发生的环境。测量效度涉及:其内容是全面的吗?是否好的表现能预测后来的成功?高低分之间是否存在有意义的差异?设计效度涉及:多久进行一次调查?何时?怎样选择被调查者?

因为调查者是在一个不完美的世界中工作,他们仔细设计的研究经常受到他们没有控制的因素的威胁。有些人说他们将参加调查,但实际没有或中途退出。研究者希望的是真实验但最终可能做不到。还有许多其他因素,这些"威胁"会导致结果的偏倚或误导。

坎贝尔和斯坦利(Campbell and Stanley,1963)识别和总结了对研究设计的威胁。他将这些威胁划分为对内部效度产生不利影响的和对外部效度产生不利影响的。如果所研究的后果(比如说,对学校态度的改善)是由研究者在研究中控制或操纵的变量(比如一个实验性项目)引起的,研究设计就是具有内部效度的。如果研究发现适用于其他人或其他地方,就是具有外部效度的。

内部效度

这里是坎贝尔和斯坦利收集的对内部效度的威胁的综述。

对内部效度的威胁	定义和例子
参与者的选择	偏倚来自选择的两组被调查者在年龄、动机、教育等方面不对等。例如,一个被调查者组主要是男性,而另一个组主要是女性。如果组不是随机选择的或随机化过程未起作用,就会出现这种情况。

续表

对内部效度的威胁	定义和例子
历史	调查期间发生的出乎意料的事件。例如,对一个基于学校的鼓励健康饮食的项目的满意度,可能受到一个流行的儿童电视节目发起的健康饮食活动的影响。
成熟	参与者内部随时间而发生的不可避免的过程(如身体和心理的成长)。例如,孩子在一个3年期的学校体育项目中的逐渐成熟。
测验	参加一个调查会影响随后的另一个调查的回答和结果。例如,在一个为期3个月的项目结束后,对参与者进行了调查。他们回忆起他们在第一个调查中的回答,而这会影响他们对第二个调查的回答。
工具	由于调查工具的变化或评分者或执行者的变化而产生的一种效应。例如,调查者甲在3次调查中对被调查者的提问方式有轻微的变化;调查者A执行问卷并负责头两次的调查,而调查者B负责第3次的调查。
统计回归	基于极值分选择的被调查者在那个变量上回归或重返到均值(如平均分)的一种效应。例如,只调查了风险最大的人,而他们中的一些人不可避免地"回归"到均值或平均"分数"。回归均值是一种统计假象(即,由于某些外在于研究的单个或多个因素使变量值变化)。
流失	被调查者从一个或多个组非随机地差别流失。例如,在参加了第一次调查后,繁忙的人可能会缺席后续调查。

外部效度

对外部效度的绝大多数威胁是由于被调查者被选择和指派到研究组中的方式。例如,在实验情境中被调查者可能以异常的方式回答问题,因为他们知道他们处于一项特殊的实验中,这被称为霍桑效应。只要被调查者被测验、调查或观察就会威胁到外部效度。他们会对被期许或赞扬的行为保持敏感。以下是坎贝尔和斯坦利收集的对外部效度的一些威胁。

对外部效度的威胁	定 义	解 释
选择偏倚和实验处理的交互效应	这种威胁发生在当一个干预或项目和参与者构成一种独特的组合(别处没有)时。当组不是随机构成的时候,这种威胁最明显。	假设一个大公司的志愿者参加一个旨在提升员工业余活动质量的实验性项目。公司的特征(其中有些,如领导方法和优先考虑,与其开展实验的意愿相关)对项目有影响,两者相交互就成为独特的,一种公司和项目的特殊组合会限制研究发现的应用。
测验的反馈效应	在干预或项目之前进行调查带来的特殊效应而产生的偏倚。	假设两组学生参加一个研究。组1在观看一部电影前进行一次调查,而组2只观看电影。组1在观影后对两组的调查中表现更为积极。第一次调查使他们对电影的内容敏感起来。
实验安排的反馈效应或霍桑效应	当被调查者知道他们参加的是一个实验而发生的效应。	有时被称为霍桑效应,这一威胁的原因是当人们意识到他们所处的环境不同时会以非常态的方式行事。(例如,当他们在教室中被摄影机拍摄,或者他们被"选中"参加一个实验时。)
多项目干扰	当被调查者参与其他补充性活动或项目时产生的干扰效应。	假设一个实验性数学项目的参与者也参加了一个物理班。两边的老师以完全相同的方式教授微分学。

调查、研究设计以及内部和外部效度

当你选择和实施一种调查设计,你几乎总是会自动地受到对内部效度的威胁。除非是完美或真正的实验,被调查者是随机选择和指派的,参与者不知道他们是在实验组还是对照组(双盲随机对照试验)。完美的真实验需要大量资源。虽然它们被认为是许多学术性研究的黄金标准,但是他们实施起来困难且成本昂贵,所以你可能找不到很多。

看看下面这个截面调查的例子,这是一种最典型的独立调查。

■ 示例

截面调查和对效度的威胁

一位调查者被要求描述医生花费在老年病人上的时间,不包括办公室访问。调查者在随机选择的 3 个 1 周的时段内,调查了一个老年医学院的 16 名医生。通过使用结构化表格(自填问卷),研究者发现医生在办公室访问之外花费了大量时间为病人提供慢性病方面的服务。调查者指出,数据是自我报告的(测量效度),并且是从一个老年医学院获取的(内部效度:选择),这些会限制研究的效度。

如果你希望更多地了解这个调查,参见 Farber, J., Siu, A., &Bloom, P.(2007).How much time do physicians spend providing care outside of office vieits? *Annals of Internal Medicine*, *147*(10),693-698。

队列设计和对效度的威胁

大雾山研究包括一个美国东南部的年龄在 9 到 16 岁的儿童和青少年的代表性样本。从 1993 年到 2000 年,这项研究中的孩子及其父母被多次访谈,了解其最近的精神病史和自杀史。访谈的目的是概括与自杀相关的风险,包括焦虑、抑郁、破坏行为和药物滥用,并检验是否存在失序的临界组合,可以在不依赖精神病理学的严重程度的情况下区分高危青年。

访谈人是研究区域内的居民并且至少有本科学历。他们接受了 1 个月的训练和经常性的质量监控。访谈通常在家中进行。调查者发现相关症状减少的程度和总的症状数量解释了自杀风险的大部分。只有抑郁加焦虑失序可以不依赖精神病理学的严重程度区分高危青年。

调查者注意到样本是在美国东南部一个相对较小的地理区域中抽取的,可能对其他地区缺乏代表性(内部效度:选择)。另外,他们没有从医学上核实报告出来的自杀企图(测量误差:同时效度)。

如果你希望更多地了解这个调查,参见 Foley, D. L., Goldston, D. B., Costello, E. J., & Angold, A.(2006). Proximal psychiatric risk factors for suicidality in youth: The Great Smoky Mountains Study. *Archives of General Psychiatry*, *63*(9), 1017-1024.

对照组设计(真实验)和效度

人们一旦遭受人际暴力就经常会再次成为受害者,再次发生率从 6%到 44%不等。这项研究假设如果满足青年的未被满足的社会心理需求,就可以减少这种高的复发率。同意参加的青年及其监护人被随机指派到一个实

验组或对照组。开发了若干个调查问卷和量表,包括养育实践问卷、家庭关系量表和孩童不法行为自我报告量表等。

调查者注意到对研究效度的许多威胁局限了研究发现的可信性。例如,研究设计不是盲化的,因而参与者可能表现出非常态行为(霍桑效应)。较大调查的几乎所有部分都经过验证,但组合的没有(测量效度)。

研究受到那些被排除在外、失去跟踪(内部效度:流失、选择)或档案不完整的青年样本的限制(测量效度)。

如果你希望更多地了解这项研究及其调查,参见 Zun, L., Downey, L., & Rosen, J. (2006). The effectiveness of an ED-based violence prevention program . *American Journal of Emergency Medicine*, *24*(1),8-13. ■

小 结

一个调查的研究设计指的是其开展的频度(一次或更多次)、其时序(新项目刚刚开始和结束)和分组数(一个学区的所有学生或者 10 个学区每一个的所有学生)以及如何被指派到组(随机或方便)。

截面调查提供描述事物目前状态的数据,并且是相对容易施行的。如果事物快速变化,调查信息可能会过时。

队列提供给调查者效果是否持续的数据。除非被调查者受到鼓励,否则他们会退出,而留下来的可能不能代表整个群体。另外,如果他们是愿意和有时间参与项目和完成调查的,也会使组不典型。

实验的特征是随机选择和指派。许多研究者断言最好的真实验是那些参加者和调查者都不知道组别的实验。

在调查的"真实"世界和社会研究中开展真实验可能是困难的。它们的成本相对昂贵,且执行和分析需要时间。然而,如果你希望知道实验是否有效果,真实验或随机对照试验才是正途。

准实验设计易于执行。然而,因为组不是随机构成的,它们最初的差异可能是引发后果的原因。

如果你能找到合适的数据库,常模调查设计较其他设计更为便宜和省时。

有效的调查信息来自可信和有效的调查工具(测量效度)和调查设计的有效性(设计效度)。

测量和设计效度的不同之处在于前者指的是调查工具的特征,而后者则是指调查发生的环境。测量效度涉及:其内容是否全面?好的表现是否能预测后来的成功?高低分之间是否存在有意义的差别?设计效度涉及:调查多久开展一次?何时?如何选择被调查者?

如果所研究的后果是由研究者在研究中控制或操纵的变量引起的,研究设计就是具有内部效度的。如果研究发现适用于其他人或其他地方,就是外部有效的。

对内部效度的威胁包括参与者的选择、历史、成熟、测验、工具和统计回归。

对外部效度的威胁包括选择偏倚和实验处理的交互效应,测验的反馈效应、实验安排的反馈效应或霍桑效应以及多项目干扰。

思考一下

1.找出后面的文献(或任意3篇包含调查的文章),说出该调查所使用的研究设计,讨论该设计的内部和外部效度。

2.解释以下术语在谈及调查时的含义:研究设计、测量效度、设计效度、内部效度、外部效度、截面调查、纵向调查、队列设计、个案控制设计、真实验、随机对照试验和准实验。

文 献

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.

Gardner, F., Burton, J., & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: Outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry*, 47(11), 1123-1132.

Patrick, K., Calfas, K. J., Norman, G. J., Zabinski, M. F., Sallis, J. F., Rupp, J., et al. (2006). Randomized controlled trial of a primary care and home-based intervention for physical activity and nutrition behaviors: PACE+ for adolescents. *Archives of Pediatric Adolescent Medicine*, *160*(2), 128-136.

Stein, B. D., Jaycox, L. H., Kataoka, S. H., Wong, M., Tu, W., Elliott, M. N., et al. (2003). A mental health intervention for schoolchildren exposed to violence: A randomized controlled trial. *Journal of the American Medical Association*, 290(5), 603-611.



6

分析和组织调查数据



分析数据意味着用统计和定性的方法描述和解释被调查者对调查问题的回答。

调查者几乎总是会报告描述性统计值，诸如被调查者的平均年龄、教育水平和年级或分数。有时候他们用列联表总结数据，以描述对两个不同的调查问题的回答之间的关系。假设一个调查者想要描述100个男孩和100个女孩对一个学校体育项目的满意度。结果的列联表形式如下：

性　别	满意的百分比/%	不满意的百分比/%
男孩	42	58
女孩	55	45

上表显示55%的女孩满意，只有42%的男孩满意。13个百分点的差异是否有意义？找到答案的一种方式是使用某种统计检验，如卡方检验。当数据是百分比时，卡方检验可用于评估两组或多组之间统计上显著的差异。如果数据是连续的，调查者可以使用其他统计检验评估显著性。其中之一是 t 检验。选择统计检验的前提是了解数据是分类的还是连续的。分类和连续数据取决于与每个问题相应的回答

尺度的类型。

统计方法允许调查者做出关于统计显著性的陈述。统计显著性是对调查结果的有意义性的度量。两组之间平均分的差异是统计显著的,意味着两组之间确实存在差异,并且更重要的是,这种差异是“真实的”,因为它不大可能是因为偶然而出现的。但是调查者也关心研究结论对于调查结果的使用者而言是否也有意义。置信区间提供了一个值域,有助于确定差异是否有实际意义。

除了描述回答和回答的差异,调查者还关心问题答案之间的关系或相关。例如,身高和体重之间的高相关,说明身高高的人体重更大以及体重大的人身高更高。回归是一种以相关为基础的用一个变量的值预测另一个变量的值的统计技术。例如,假设你希望了解中学阶段参加体育运动是否与大学参加体育运动相关。为了找到答案,你可以对调查数据做回归分析。

调查者使用的其他技术还包括风险和几率。这些统计技术提供了某组内发生某一特定结果的可能性,也可以用于对结果的组间比较。

内容分析是一种能够用于找出开放性调查问题的回答的共同观点或话题的定性方法。内容分析需要对开放问题的回答及评论编码。如果足够量的人不断地写他们关注相同的问题或表达某一个观点,这就成为一个话题。

数据管理是数据分析的一个有机组成部分。管理数据意味着把数据组织为变量并进行命名和编码,以便一旦将数据录入数据库就可以对它们进行分析。对数据进行编码是一种基本的分析活动。在大型调查中,你必须检验编码的可靠性。通常使用两名编码员。为了确定编码是否可靠,可以计算一个称作 Kappa 的统计量。Kappa 提供了一个对两个编码员之间的一致程度的度量。

你必须检查调查数据中的缺失值并决定如何处理它们。缺失值来自对某个问题或整个调查的未回答。

数据录入是将调查的回答输入数据库的过程。网络调查的数据已经自动地输入数据库了,并且能实时地获得报告。调查数据还可以通过扫描或手工输入电子表再导入数据库。无论数据是怎样输入的,

都需要首先对回答的正确性进行检查。例如,调查者通常不希望从男人那里获得关于生育时的生理快感的回答。大多数网络调查程序都提供了自动检验正确性的选项。一个"干净"的数据集指的是任何人使用它都可以得到与你相同的结论。数据变"脏"有多种原因,包括不好的问题、误导性的回答、错误编码、不正确的数据录入以及缺失回答。为了避免"脏"数据,需要确保编码员或数据输入者是有经验的、受过训练的和受到监督的。如果你在使用网络调查软件,你能找到"验证"功能使回答规范,具有一致性并符合你的标准。这有助于获得一个"干净"的数据集。

什么是常见的?分析调查数据的一些常用方法

这一天到来了。所有的调查都已经收回。回答率是高的。所有的问题和所有的调查表都完成了。现在是时候找出调查结果了:有多少男人作答?有多少女人作答?他们有差别吗?他们的看法在过去的5年里是否发生了变化?调查者通过分析调查数据,获得回答频数和平均回答来找出关系、差异和变化,以此回答诸如此类的问题。下面是一些常用的分析方法。

描述性统计

它们是描述一个调查的被访者和回答的基本单位。描述性统计提供关于样本和对某些或所有问题回答的简单汇总。与简单的图形分析(如饼图和条形图)一起,共同构成了各种调查数据量化分析的基础。调查的描述性统计包括频数或频数分布(数量和百分比)、集中趋势的度量(均值、中位值和众数)和离差的度量(极差和标准差)。

频数。频数是对某一类中有多少人(如20人超过55岁;150人患感冒)或选择某一回答的人(如,32人说极其重要;8个人说每周超过3次)的计数。

考虑这个:调查了50名幼儿园园长。所有的幼儿园都是公立的。调查见后。

■ 示例 幼儿园目标调查表

以下是一些可能的幼儿园教育目标。针对每一项圈选对你的幼儿园教学的重要程度。(数字 1 为非常重要,数字 5 为非常不重要。)

目　标					
培养阅读、数学和科学方面的成绩	1	2	3	4	5
通过音乐、舞蹈和艺术鼓励创造性	1	2	3	4	5
促进公民教育	1	2	3	4	5
增进社会和个人发展	1	2	3	4	5

■

对于幼儿园目标的问题,你可以按下例所示汇总回答。

■ 示例 问卷回答的频数

选择某一目标的幼儿园园长的数量和百分比。

目　标	*N*	%
学业成绩	20	40
创造性	13	26
公民教育	11	22
社会和个人发展	6	12

■

有时候你希望把回答组合在一起(或者用技术术语来说,提供一组回答的频数分布),如下例所示。

■ 示例 50 名园长对幼儿园目标的组合评价

	选择这一目标的园长人数	
	非常重要或重要	非常不重要或不重要
目　标	*N*	*N*
学业成绩	40	0
创造性	15	30
公民教育	26	5
社会和个人发展	7	42
合　计	88	77

■

在这张表中,回答被分为两类(离散的数据):重要(包括两类回答:非常重要和重要)和不重要(包括两类回答:非常不重要和不重要)。为什么要把回答组合在一起?因为如果仅有少数人选择了某个选项(比如只有 3 个非常重要的回答),那么该类别对你就失去了意义。

平均数:均值、中位值和众数

均值、中位值和众数都是对平均或典型情况的测度。

均值。算数平均值,均值,需要把每一份加总起来再除以总份数。这里有 5 个分数:50,30,24,10 和 6。这 5 个分数的平均值是(50+30+24+10+6)÷5=24。

均值的公式是:

$$\mu_x = \frac{\sum x}{N}$$

式中 μ 是均值的符号;$\sum$ 表示连续加,所以 $\sum x$ 表示把各个 x 的数值加在一起;N 代表 x 的数量。

假设你想计算四个幼儿园目标的平均评级。首先,你需要知道园长们赋予每个目标的评级。见下表:

50 名园长的评级

	1.非常重要	2.重要	3.中等	4.不重要	5.非常不重要
学业成绩	26	20	4	0	0
创造性	13	2	5	20	10
公民教育	11	24	10	10	0
社会和个人发展	6	1	1	30	12

为了计算每个目标的平均评级,你需要将选择每个刻度的园长人数乘以该刻度的数值。对于创造性,13 名园长选择了 1,所以就是 13×1,把结果加在一起(13×1)+(2×2)+(5×3)+(20×4)+(10×5),然后除以园长人数:(13+4+15+80+50)÷50=3.24,这就是对创造性的平均评级。

在这个案例中,平均值越接近 1,该目标的重要性就越高。看下表:

50 名园长做出的平均重要性评级

目　标	平均值
学业成绩	1.56
创造性	2.18
公民教育	3.24
社会和个人发展	3.82

中位值。中位值是等级中的一个点,各有半数的分数高于和低于此值。换一种说法,中位值是 50 百分位。因为中位值总是落在中间,所以当你希望描述"典型的"表现时可以使用它。

为什么你需要了解典型的表现?假设你有一组分数类似于:5,5,6,6,6,8 和 104。平均值是 20,而中位值是 6。为什么会发生这种情况?因为存在极端值。当你抽选出来的组在态度(或者知识、健康等)方面有分化时,大多数人是一种感觉,而某些人的感觉非常不同,就会发生这种情况。当你未能收集到你计划收集的所有数据,具有某种观点的人在回答中没有充分代表时,也会出现这种情况。

以下是当你有偶数个分数时如何计算中位值:

1.将分数按大小排序。

2.中位值处于第 $N/2$ 个分数和 $(N/2)+1$ 个分数(其中 N 为分数的数量)之间,使用算数平均值。

■ 示例　　**偶数个分数如何计算中位值**

有这样一组分数:-2,0,6,7,9 和 9。

共有 6 个分数,所以 $N=6$,而 $N/2=3$ 以及 $(N/2)+1=4$。按顺序第 3 个分数为 6,第 4 个分数为 7,所以中位值是 6.5。

再看这组分数:2,4,5,8,9 和 11。

同样,$N=6$,所以中位值位于第 3 和第 4 个分数之间。然而,这一次第 3 和第 4 个分数之间相差 3 个单位。把 2 个分数加起来 5+8,再除以 2,得到中位值为 6.5.

如果一组数据不多且为偶数,可以这样做:

1.把分数按大小排序。

2.中位值为距离分布底部(N+1)/2 处的分数。

■ 示例　　**奇数个分数如何计算中位值**

看下面这组数字:-9,-8,-6,-6,-4,-3,-2,0 和 2。

这 9 个分数的中位值是(9+1)/2 或第 5 个分数,所以中位值是-4。

但是假如无论奇数还是偶数,在中位值或相邻处有几个相同的分数,例如这个数字集:3,6,7,8,8,8,9,9,10 和 12。

当 N=10 时,就像这个例子中,中位值通常落在第 5 和第 6 个分数之间。但有些学者认为本例中把 8 作为中位值是不妥的,因为你需要 5 个不同的分数在上,5 个不同的分数在下。

但是如果你把 8 分作为连续谱的一部分,均匀分布在 7.5 到 8.5 区间内,就可以通过插值获得一个中间分数。

这样考虑:7.5 到 8.5 的区间(其中包含 3 个 8)之前有 3 个个案(3,6,7)。在这个区间内,你需要增加两个个案以使总数达到 5。所以,你增加 2/3 个距离,或者说 0.67 到 7.5 上,因而得到一个 8.17 的中位值。

众数。众数是一个分数(或者一个分数等级中的点),其出现的频数高于邻近的其他分数。看下面这些分数:

分布 A		分布 B	
分　数	频　数	分　数	频　数
34	2	34	0
33	6	33	1
32	8	32	7
31	11	31	11
30	15	30	4
29	18	29	3
28	16	28	7

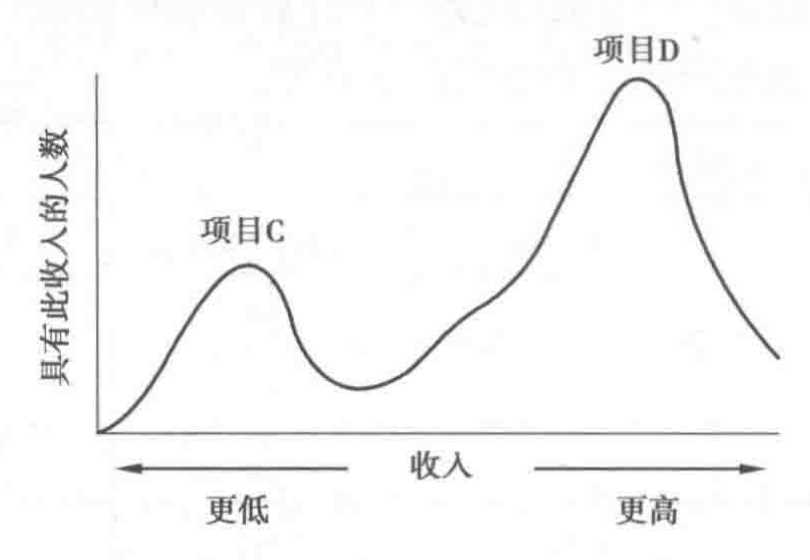
续表

分布 A		分布 B	
分　数	频　数	分　数	频　数
27	12	27	10
26	8	26	18
25	3	25	123
24	1	24	11
23	0	23	5

分布 A 有一个唯一的众数 29,有 18 个人得到此分数。这种分布称作单峰的。分布 B 有两个众数,分别是 25 和 31,因而这种分布称作双峰的。(虽然 31 分的 11 人等于或低于其他分数,但这个分数还是个众数,因为它出现的频数高于附近的其他分数。)

众数,作为集中的分数,描述了占优势的观点。当你怀疑群组具有很多不同的特征时可以使用众数。例如,如果你怀疑人们在项目 D 中比在项目 C 中做得更好的原因在于开始时的经济条件更好,你就可以比较他们在参与每个项目前的收入。

如果你的发现类似下图这样,你的观点就可能是对的。



项目 D 中的人更多在高收入的部分,而项目 C 中的人在低收入部分。合在一起,C 组和 D 组形成了一个双峰分布。

离差：极差、方差和标准差

当你计算算数平均值时，每个数字都参与计算。如果 1 或 2 个人的得分非常高或低，就会虚假地提高或降低平均分。使用中位值要好一些，但也可能是误导性的。考虑下面这组分数：2，3，4，5，6，7 和 8。

均值和中位值是 5。如果你把最后两个数字改为 17 和 18，中位值仍然是 5，各有 3 个个案高于和低于这个分数。均值将提高到55/7=7.86。

如果每个数字都参与计算，有时候很重要的一点是研究一下数字的分布即它们的离差，以此揭示平均值的含义到底为何。是否所有的分数都接近算数平均值？是否有些非常高或者非常低？看下表：

60 个继续教育课程的参与情况

学校编号	教师人数	参与课程的平均数量	极　差
1	50	55	16-60
2	103	54	53-60
3	86	41	15-59
4	117	47.5	15-60

如果你看第 4 所学校的参与课程的平均数量，会发现是 47.5。进一步看，在有些学校，只参与了 15 个课程；而在另外一些学校，所有 60 个都参与了。实际上，很难不注意到参与的极差范围。因为参考了极差，你至少可以得出两个推论：

一、4 所学校中的 3 所（1，3 和 4）存在离差，有些教师参加了很多课程而其他人则参加得很少。

二、只有 1 所学校（2）几乎所有人都参加了和别人差不多数量的课程。

在某些情况中，离差被认为是一种有用的东西。一个训练人们独立地或创造性地思考的项目可能希望调查能够揭示多种观点。你也需要离差以做出比较。如果每个人都做得同样好或分享相似的观点，你将不能选出最好的、最强的、最自由的或最保守的，等等。然而，在

另外一些情况中,你希望每个人都能分享相似的观点,或者以同样的熟练程度获得某种技能,这时离差又是不被期望的。如果开展继续教育的学区相信他们的项目是有价值的,那么宽范围的参与率是不被期望的。

另外一个离差的测量称作方差,其平方根称为标准差。后者是一个统计术语,基于分数与平均值的距离得出。实际上,标准差是平均分与均值之间的平均距离。有时候,作为标准差的替代,使用方差。方差只是标准差的平方。

你可以只对类似平均分这样的连续数据计算方差和标准差。研究分类数据(包括分类和顺序尺度)时则使用极差。

相关和回归

数据是相关的意味着它们是有关联的。例如,身高和体重之间高的相关,意味着较高的人体重更大以及较重的人身高较高。相关系数的范围在+1(高度正相关)和-1(高度负相关)之间。

当一个变量的高值与另一个变量的高值同时发生时,就说这两个变量是正相关的;当一个变量的高值与另一个变量的低值同时出现时,就说两个变量是负相关的。以 r 代表相关系数,通常保留两位小数。

对于分类数据,也就是说来自分类或顺序尺度的数据,使用斯皮尔曼等级次序相关系数。等级次序相关系数提供了两个等级序列之间相关或对等程度的度量,如下例所示:

■ 示例　　**等级次序相关**

一个有 50 名学生的大学班级接受了两次态度调查。第一次调查了解他们对一般保健的观点,第二次调查询问他们的政治偏向。约翰在第一次测量中得分最高(态度最好),在第二次测量中处于中游;简的得分分别处于从高到低的第 5 和第 8 位;比尔得分处在第 14 和第 13 位,等等。可以计算等级次序相关系数来看一看两个调查是否一致:在一次调查中处于高位的人是否在另一次调查中也处于高位? ■

如果调查产生的是像身高和体重这样的连续数据,使用何种统计

程序合适呢？皮尔森积矩相关系数用来建立两组连续数据之间的关系。以下是适合使用此类相关系数的两种情形：

• 平均成绩（GPAs）与对学校态度调查中的得分之间的相关关系。

• 自由主义和保守主义观点（1=自由主义到 10=保守主义）和家庭收入之间的相关关系。

回归是以相关为基础通过一个变量的值预测另外一个变量的值。例如，假设你希望知道学生在中学参与体育运动是否关联（预测）其在大学参与体育运动。为了找出答案，你可以对调查数据采用回归分析。

再举一个例子，假设你希望了解中学成绩好的学生是否大学成绩也好。换句话说，你希望知道中学成绩好是否能够“预测”大学成绩好。为了找出答案，你首先调查大学新生以获取他们中学和大学一年级时的平均成绩。然后你可以考察这两个成绩之间的相关性。最后，你可以重新调查一组中学生（接近毕业尚未进入大学的），然后使用回归技术，基于之前的相关性“预测”他们大学一年级的成绩。

警示：你可以使用相关去识别变量之间的关系，但是你不能用他们建立因果关系。假设你访谈了一组人了解他们的最高学历（变量1）。然后再问他们今年订阅杂志的数量（变量 2）。对教育水平和杂志订阅之间相关性的分析显示样本中受教育年限越长的人订阅杂志的数量也越多，但这并不能解释为受了多少年教育决定了订阅多少杂志。可能是样本中受教育年限长的人也更富有，使他们能够订阅更多杂志（能够负担得起），还可能是你的样本不够典型或者太小，以致不能得出结论。

多元回归技术能够让你考虑不止一个预测因子。假设你综述了教育文献，发现中学平均成绩高的学生来自家庭收入相对更高的家庭。也就是说，文献支持家庭收入和学生的中学平均成绩之间存在高相关的假设。你现在可以提出这样的问题：中学成绩（预测因子 1）和家庭收入（预测因子 2）对大学平均成绩的贡献各是多少？一个多元回归分析能够帮助你回答这个问题。

组间差异

描述性差异:列联表。列联表(经常简写为 cross-tab)是用来同时描述两个变量的一种方法。它经常以表格的形式出现。表格的名字通常是根据行列数命名的。

下面是一个假想的3(行)×2(列)列联表的例子,描述了一个旨在了解各种电子通讯方式使用的调查结果。在这个例子中,调查者创建了一个表用来描述对两个调查问题的回答:1.在典型的一天中,你花费在网上的时间有多少?回答选项是少于1小时、1~3小时和超过3小时。2.你对调查的感觉如何?回答选项是喜欢调查和讨厌调查。表格中的每个单元格给出了具有该组合特征的被调查者的百分比。例如,在这个调查中,喜欢调查者中有70%的人每天上网时间在1小时或以下。

列联表易于理解,并且能为理解调查结果的含义提供洞察力,因而经常被使用。从这张表中可以看到,讨厌调查的人有60%每天上网时间超过3小时。对于打算开展网上调查的人来说,这可能会是一个有趣的发现。

	你对调查的感觉如何?	
每日上网时间	喜欢调查	讨厌调查
少于1小时	70%	5%
1~3小时	25%	35%
超过3小时	5%	60%

现在考虑这个。一家市场研究公司希望调查MacWindow公司的顾客,以了解顾客们有可能在接下来的12个月内购买何种产品。这一信息被用来规划公司的库存。也就是说,MacWindow公司将根据多数顾客的购买计划备货。MacWindow公司开展了一项网上调查以了解其顾客的计划。调查的一部分复制在下面:

1. WHAT DO OUR CUSTOMERS PLAN TO PURCHASE?

1. Do you plan to purchase any of these in the next 12 months?

	Yes	No	Don't know/Not applicable to me
Notebook computer	○	○	○
Desktop computer	○	○	○
Digital camera	○	○	○
Video camcorder	○	○	○
DVD or Blue Ray player	○	○	○
Cell phone	○	○	○
Two-way radio	○	○	○
In line skates	○	○	○
Snowboard	○	○	○
Skis	○	○	○
Bicycle	○	○	○
Video game system	○	○	○

2. What is your age?

Your Age

Please select one.

3. What was your annual household income in the past calendar year?

- ○ $25,000 or under
- ○ $26,000 to $50,000
- ○ $51,000 to $75,000
- ○ $76,000 to $100,000
- ○ $101,000 to $200,000
- ○ $201,000 or more
- ○ Prefer not to state

4. Are you male or female?

- ○ Female
- ○ Male

调查者给出了两个问题的结果的一个列联表:“你未来 12 个月内打算购买何种产品?”和“你的年龄?”

	未来 12 个月内是否打算购买下列产品? 单元格中的数字是回答“是”的百分比。被调查者可以多项选择。												
你多大年龄	笔记本	台式机	视频游戏系统	数码相机	摄像机	DVD/蓝光播放器	蜂窝/移动电话	无线对讲机	直排轮滑鞋	滑雪板	滑雪橇	自行车	被调查者合计数
<19	—	—	36	12	—	—	—	—	12	12	25	50	8
19~23	5	2	19	7	10	12	17	12	29	17	19	31	42
24~29	5	4	19	10	17	12	15	10	15	30	11	30	353
30~35	8	4	20	10	23	11	8	9	12	32	9	28	192

续表

你多大年龄?	笔记本	台式机	视频游戏系统	数码相机	摄像机	DVD/蓝光播放器	蜂窝/移动电话	无线对讲机	直排轮滑鞋	滑雪板	滑雪橇	自行车	被调查者合计数
36~45	6	7	19	5	15	17	12	12	17	30	5	36	165
>45	8	4	12	8	18	24	16	9	16	31	2	36	129

从列联表中的数据,MacWindow 公司可以看到某些模式。例如,最大数量的被调查者集中于 24~29 岁,有大约 1/3 的人打算在未来 12 个月内购买自行车。这是一个相当大比例的潜在顾客群,因而 MacWindow 公司决定多备各种自行车以满足这些顾客的需求。第二大的被调查者群体介于 30 到 35 岁之间。这个群体中有 23%的人打算购买摄像机,32%的人打算买滑雪板。实际上,在所有年龄超过 30 岁的被调查者中几乎有 1/3 的人计划在未来 12 个月内购买滑雪板。

MacWindow 公司还可以通过其他列联表回答公司关心的其他问题。例如,公司决定考察一下年龄和收入之间的关系。是否收入随着年龄而提高,或者是否他们的年轻顾客比年长顾客更多地来自更为富裕的家庭?男性和女性的家庭收入是否有所不同?购买计划上是否有所不同?高收入家庭和低收入家庭在购买计划上是否存在差异?这些都是好问题,但不幸的是,列联表只能描述收入和购买计划。它不能告诉调查者观察到的差异在统计上是否有意义。因此,调查者必须依靠标准的统计方法。

统计差异

假设你想知道一组被调查者是否与另一组有差异?他们是否更健康?更可能在下次选举中投票?更适于受雇?拼写更好?更可能购买 MacWindow 公司的自行车或数码相机?检验组间差异的方法包括卡方检验、t 检验和方差分析(ANOVA)等统计技术。

统计检验实际上假设各组的调查结果都是相似的。例如，如果你想比较组 A 和组 B 的态度分并使用 t 检验这样的统计检验，你假设平均分是相同的。要点是建立一种严格的环境以对这个假设进行检验。

对等同性即“被调查者组是相似的”的检验被称作零假设。在陈述这一假设后，就进行统计分析，产生一个统计量，以此为指标决定接受还是拒绝零假设。如果零假设被拒绝，备择假设（组是不同的）得到支持，那么结果就被认为是统计上显著的。

假设你调查一艘游轮上的人，其中半数参加了一个把旅游和教育结合在一起的实验性项目。你十分肯定这个项目是有用的，实验组的满意度远高于对照组的满意度。然而，作为一个严谨的调查者，你需要从两组的满意度相同的假设开始。然后，你使用一种称作独立样本 t 检验的统计方法比较两组的平均满意度得分。你发现你是对的，所以你拒绝没有差异的假设。毕竟存在统计上显著的差异。因为实验组的平均满意度分数更高，你得出结论实验有效，你最初的想法是正确的，而且现在有统计数字支持你。你结论的关键点是统计显著性的概念。如果结果不太可能是由于偶然而发生的，它们就是统计上显著的。

统计显著性

假设你调查两组学生的态度，其中一组参与了一个实验性的阅读项目。再假设实验组的分数低于其他组——比如说，10 分。这相对较低的分数是由于偶然，还是因为新的阅读项目？不可能由于偶然而发生的结果称为统计上显著的。两组差异需要达到多大才能排除偶然因素的影响？

为了确定统计显著性，你必须依靠抽样理论。例如，你提出一个这样的问题：来自同一总体的两个随机学生样本的平均分相差 10 分或者 20 分的可能性有多大？

假设你设定 100 次中由于偶然而发生 1 次是可以接受的。这个预设概率（$p<0.01$）被称作显著性水平。如果你观察到的差异在 100 次中发生的次数不足 1 次，就可以拒绝组间无差异的零假设。

按照惯例，调查者通常使用 0.05，0.01，0.001 等显著性水平，这意

味着如果10分的差异由于偶然而发生的几率等于或小于100次中发生5次、100次中发生1次或者1 000次中发生1次,那么就可以认为观察到的实验和传统项目(假设两组是从同一总体中随机抽样)的差异是统计上显著的。

在检验统计假设时,你必须建立何时接受或拒绝零假设的规则。举个例子,对一个实验(A)和对照(B)阅读项目的统计检验,零假设是两组的平均阅读分数是相同的。当你对数据应用统计检验(如t检验)时,你不希望发现两组之间的平均分"零差异"。相反,真实的问题是差异是否小到是由于偶然而发生的。当你从一个总体中选取两个随机样本,你可以预期他们的平均分是接近的,但不会完全相同。

需要由你来决定分数必须相差多少才能认定差异不是因为偶然而产生的。你可以根据你能容忍的拒绝零假设所犯错误的数量,选择0.10,0.05,0.01或0.001作为显著的标准。为什么是这些数字?因为它们是调查者和研究分析者根据传统而惯常使用的。

如果你选择0.05的显著性标准,那么100次可能有5次会拒绝其实为真的零假设。这种情况之所以会发生是因为当你比较两个来自同一总体的随机样本时,仅仅是因为偶然而产生差异的概率是5%($p<0.05$)。这种情况被称为I类错误。它是拒绝为真的零假设的概率。也就是说,当你说存在差异,而事实上不存在差异时,你就犯了一个I类错误。

如果显著性水平是0.01,那么犯I类错误的概率100次中只有1次,或者说1%。你可以选择一个很高地排除犯I类错误机会的显著性水平,但会产生严重的后果。你犯I错误的可能性越小,犯Ⅱ类错误的可能性就越大。

Ⅱ类错误是你接受了一个实际上存在的零假设。例如,你认为两组之间不存在差异,而实际上存在差异。当你犯Ⅱ类错误时,两组的平均分的差异没有落在拒绝区域内(比如说,$p<0.05$)。

一个统计检验的效力指的是该检验拒绝一个错误的零假设(将不会犯II类错误)的可能性。另外一种考虑方式是检验检测出真实存在的给定效应的能力。随着效力的增加,犯Ⅱ类错误的机会减少。Ⅱ类错误的概率指的是假阴性率(β)。因此,效力等于1-β。

效力分析可以在收集调查数据之前(事前地)或之后(事后地)。事前的效力分析通常用于确定一个适当的样本量以获得充分的效力。事后的效力分析使用已获得的样本量和效应量确定研究中的效力如何。

效应量帮助确定一个统计上显著的差异是否具有实际上的意义。这是一个通用术语,指的是对所研究的效应的估计,通常界定为实验组与对照组的均值差再除以对照组或两组的标准差。如果样本量足够大,所比较的两个均值之间几乎总是会存在一定差异。效应量帮助了解所观察到的差异是否有意义。

对分析组间差异感兴趣的调查者通常用置信区间(CIs)来表达这种差异,而不是 p 值。CI 是对均值这样的总体参数的估计,但却被描述为在一个区间或范围内。

假设你调查了在一所大学中随机选择的 300 名学生。你问他们上个月在网络购物上花了多少钱。样本的平均值是 100 美元/月。但是有可能这些学生的网上花费要比该所大学的其他学生少。例如,总体的均值(这所大学所有学生的均值)可能接近 500 美元。少数大学生可能每月花费数千美元,而这会大大提高总体的均值。

样本均值是调查者无法绝对确定的(除非样本就是整个总体)。样本量越大,调查者对样本均值"相当接近"总体均值的把握就越大。

CI 之所以有用是因为其基于重复抽样的理论。如果反复从一个总体中抽取同样大小的样本,那么这些区间将以某个比例——比如说 95%——包含总体的均值。

参看下表:

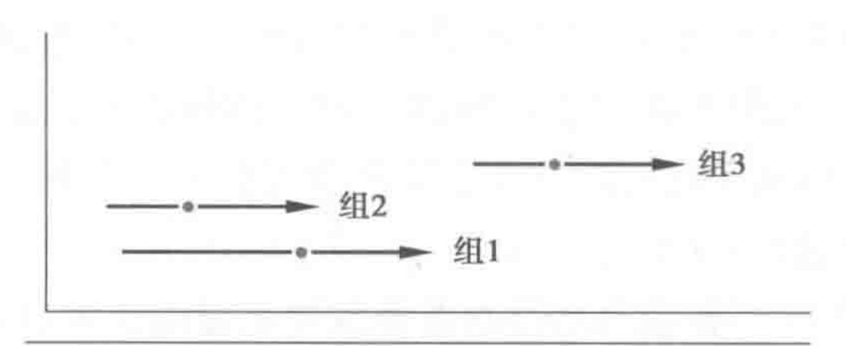
组	均 值	标准差	标准误	95%CI(均值的)
1	11.00	3.60	1.36	7.67—14.33
2	8.14	4.30	1.62	4.17—12.11
3	16.43	4.98	1.09	13.51—19.35

该表的结果中包括一些描述性统计量:均值、标准差、标准误和 CI。组 1 的均值是 11,该分数的 CI 是 7.67—14.33。如果从组 1 来自的总体中反复抽取相同量的样本,然后对每个样本计算 CI,那么这些

区间中有95%应该包含总体均值(介于7.67到14.33之间)。

你可以把CIs画在一张图上。如果均值不重叠,那么差异存在。如果一组的均值包含在第二组的区间内,则差异不存在。如果区间而非均值重叠,那就不能确定。

参看以下的CIs图。



组1的置信区间7.67到14.33, $\overline{X}=11.00$
组2的置信区间4.17到12.11, $\overline{X}=8.14$
组3的置信区间13.51到19.35, $\overline{X}=16.43$

组2的均值在组1的置信区间内。组3的区间与其他两组都不重叠。上图中能够看到均值的差异,并且你可以拒绝零假设(即均值相同)。顺便说一句,如果你使用类似ANOVA的统计方法,你会发现p值和CI的结果是一致的。CI和p值是相关的。如果区间包含0,那么p值将不显著。

CIs一般用于描述差异。他们会被用在许多地方,因为能让你看到值的范围。

调查差异:常用方法

以下是一些评估组间差异(或者更准确地说,检验它们是否相同)的常用技术。

卡方

卡方检验用于分类数据。它检验的是比例形式的调查数据是否相等的假设。比例是所有可能的人中以某种方式作答或具有某种特征的人的多少。假设你抽取了1 000人,发现其中250人的眼睛是蓝色的。蓝色眼睛人的比例就是250/1 000。再假设这蓝眼睛的250人中有60人为女性。假设你想比较蓝眼睛的男性(190/250)和女性

(60/250)的比例。你估计这个比例是“显著不同的”(换句话说,它们是真实而非偶然的差异),可用于比较比例的统计检验就是卡方。卡方的符号是χ^2。以下是如何应用卡方的一个例子。

■ 示例　　卡方

监狱组织调查了60名男性囚犯和40名女性囚犯,以了解他们认为监狱系统是公平的还是不公平的。结果如下表:

对监狱系统的看法	男	女	合　计
公平	57	32	89
不公平	3	8	11
合　计	60	40	100

■

该表显示57名男性和32名女性认为系统是公平的,3男8女认为是不公平的。卡方可用于检验的零假设(“无差异”)是在他们关于公平性的看法上男性和女性的比例是相同的。如果零假设成立(“可被接受”),你就得出结论“无差异”。该表称作2×2表。它看上去像列联表,但这次将对数据进行统计检验。第一个“2”包含男女两列,而第二个“2”是行的公平和不公平。

t 检验

t 检验也用于检验差异。它允许你比较两组的均值,以确定两者的差异是真实而非偶然的可能性。统计教科书会告诉你只有每组至少有20~30名被调查者且为连续数据时才能使用 *t* 检验。你需要连续数据计算算数平均数或均值。

■ 示例　　*t* 检验

Hope医院启动了一个美食计划。Charity医院认为美食花费太多。两所医院的病人对膳食的满意度存在差异吗?调查者调查了Hope医院的50名病人和Charity医院的60名病人。他们使用了满意度量表,要求人们在一个10分的连续量表上打勾。调查者计算了每所医院的平均满意度,然后使用 *t* 检验比较了结果。

■

Mann-Whitney *U* 检验

Mann-Whitney *U* 检验(也称作 Wilcoxon 秩和)能够让你在无法使用 *t* 检验的时候(比如说样本量太小)对两个独立样本进行比较。这个统计方法是对中位数是否等同的检验。

■ 示例 Mann-Whitney *U* 检验

Partridge 小学有 10 名 4 年级学生有严重的听力障碍。其中 4 名学生参与了一个教他们更清楚地说话的新项目。在一个学期的末尾,将新项目的学生与传统项目的 6 名学生进行了比较。一位特教专家,通过在一个 20 个代表语音清晰的项目上打出 1~20 的分数,评估每个孩子的说话能力。如果学生更多的话,调查者可能会选用 *t* 检验比较两组,但是考虑到小样本问题,他决定使用 Mann-Whitney *U* 检验。 ■

ANOVA

组均值或平均数也可用 ANOVA 来检验。这种方法让你检验来自两组或更多组的平均数。例如,你可以检验实验学校 A 和实验学校 B 与对照学校 C 的学生在一个全国性测验中的平均分的差异。

风险和几率

风险和几率用于描述某种特定后果在组中发生的可能性,也可用于组间比较。假设每 100 个患感冒的人中有 20 个人同时还有咳嗽。感冒并伴有咳嗽的风险是 20/100 或 0.20。感冒并伴有咳嗽的几率的计算是将有咳嗽(20)和无咳嗽(80)的人数相比,或 20/80,即 0.25。

当你使用风险和几率比较组时,你比较的是一个后果将会发生的相对可能性。相对风险表达了某种特定后果发生在实验组的风险相对于该后果发生在对照组中的风险。几率比是实验组中的几率与对照组中几率之比。你可以以 2×2 表格的形式使用相对风险或几率比。参看下列表框架。

■ 示例

几率比和相对风险

后　果	实　验	对　照
仍在吸烟		
戒　烟		

■

你可以提出下列问题:实验组中仍在吸烟(或戒烟)的几率与对照组相比如何?实验组中仍在吸烟(或戒烟)的风险相较于控制组中仍在吸烟(或戒烟)的风险如何?

如果后果在实验组中较在控制组中更不经常发生的话,那么相对风险和几率比将小于1。类似的,如果后果在实验组中较在控制组中更经常发生的话,那么相对风险和几率比将大于1。

为什么要使用几率比或相对风险?为什么不仍然用卡方?答案在于置信区间。几率比和相对风险(而非卡方)也可以有置信区间,同样也是为了实现显著性检验的目的。也就是说,你可以确定(通过查看区间)结果是否在统计上显著。下表描述了使用网上健康信息的调查结果。

该表有5列5行。第一列包含人口统计学特征。因为该表展示的是几率比的结果,所以每个特征都被假定是与其相反面比较:男性比女性、已婚比未婚等。每个单元格都包含一个几率比和一个置信区间。在女性这一行和"曾经使用过网上健康信息"这一列可以找到三个信息:1.23是几率比,1.06~1.43是置信区间。通过这些数字,你可以看出女性比男性显著地更加可能(几率比大于1且置信区间不包含0)报告曾经使用过网上健康信息(当与未使用过网络相比时)。此外,更多的教育和更高的认知能力分数也会增加网络使用的几率。

使用网上健康信息的调查(N=4 528)

人口统计学特征	曾经使用过网上健康信息(N=2 123)	看病前使用过(N=305)	看病后使用过(N=645)	使用与看病无关的信息或替代看病(N=1 127)
女性	1.23(1.06,1.43)	1.34(0.99,1.81)	1.12(0.9,1.4)	1.30(1.09,1.56)
已婚	1.07(0.89,1.29)	0.86(0.61,1.22)	1.10(0.84,1.43)	1.09(0.87,1.36)

续表

人口统计学特征	曾经使用过网上健康信息($N=2\ 123$)	看病前使用过($N=305$)	看病后使用过($N=645$)	使用与看病无关的信息或替代看病($N=1\ 127$)
更高的中学认知能力分数	1.14(1.05,1.24)	1.18(1.00,1.38)	1.25(1.12,1.41)	1.09(0.99,1.2)
更多的教育	1.09(1.06,1.13)	1.12(1.05,1.19)	1.07(1.03,1.13)	1.10(1.06,1.14)

可能有很多统计方法适合于调查者的调查。假设一个调查者对比较新老项目中人们的满意度感兴趣。调查者估计参加老项目至少 3 年的人可能比参加时间较短的人的满意度低。也就是说,参与时间,称为协变量,能够影响满意度,即使它与新项目无关。调查者可以通过使用一种称作 ANCOVA 的统计方法消除协变量的效应。

数以百计的类似有用的统计技术可供调查者使用。特殊形式的 t 检验和 ANOVAs 可用于测量历时性变化。独立样本 t 检验可以测量一个组从时点 1 到时点 2 的变化。重复测量的 ANOVA 能被用来检测一个或多个组在两个或多个时点的变化。

在 McNemar 检验中,每个人作为他或她自己的对照,可用于小样本和分类数据。例如,你可以使用 McNemar 检验找出参与一个辅导项目的学生(与没参与的学生)在 20 岁后选择和未选择(分类)软件工程职业的人数上是否存在差异。

要还是不要:统计学家或定性分析者

调查者不一定要成为统计学家才能做调查。然而,他们必须充分理解统计数字和统计方法,以便寻求统计学家的帮助和正确使用网络调查商提供的功能。调查者还应该能够评估其他人在调查报告中使用或推荐的分析技术的适当性。

不是所有的调查数据都适合做常规的统计分析。有些调查信息

包含对开放性问题的回答和被调查者的评论。这些回答和评论可能或长或短。这些数据通常称作定性数据，与来自大多数调查中封闭问题的统计或定量数据有所不同。

定性数据可能是复杂的，因为研究者必须根据他或她自己对被调查者意图的理解来过滤回答。定性分析通常由两个或更多人来完成，以消除单个研究者的偏差并接近“真实”。有很多评论的大型调查可以使用定性分析软件来辅助。

内容分析、开放性回答和评论

内容分析用于分析定性数据，旨在推论出被调查者做出的开放性回答和评论的意义。对于调查者而言，这意味着仔细阅读评论，以找出某些字眼、概念、主题、短语、特征或句子的出现并予以量化。下面是对一个假想调查的结果的内容分析，目的是找出以不同形式取笑男孩和女孩的不同效果。

假想的内容分析：取笑男孩和女孩

对所有参加者提出了一系列问题，包括“你取笑了什么方面？”“你对被取笑的感觉如何？”“你如何应对被取笑？”“如果有人取笑过你的外貌你是否做出了改变？如果是的话，如何改变的？”

根据性别对回答进行了分类。每个性别内的回答根据问题序号排列（如对“你对被取笑的感觉如何？”这个问题的回答被放在一起以供分析）。对具体问题的回答中浮现出来的主题进行集中考察并进行两组间的比较。注意两组的不同以对可能的组间差异做出评估。随着内容从数据中浮现出来，被调查者对特定问题的评论被纳入进来以支持研究发现。

女孩经常报告对被取笑负面的意见和态度（41.3%），以及负面的情感（28.23%）、剩余效应（19.6%）、取笑的社会方面（8.7%）和应对取笑所采取的行动（2.2%）。

与女孩相比，男孩更经常报告取笑的社会方面。如果他们被朋友取笑的话，这种体验被描述为幽默。例如，一个被调查者说：“这取决

于他如果是我的朋友,我就认为是个玩笑,但如果不是我的朋友,我就会认为是恶意和有害的。"另外一个人说:"对于我的朋友,我会对被取笑无所谓,因为我也取笑他们。"

把马牵到车前:选择分析方法

恰当的数据分析方法取决于样本量、调查的研究设计以及数据的特点和质量。经常有人问:"分析调查数据最好的方法是什么?"回答这个问题前必须首先回答下面7个问题。

1.**你调查了多少人**? 在选择分析策略时样本量是一个重要的考虑因素。有些统计方法需要的样本量要比其他方法更大。看下面这些样本:

A)学区内的所有500名教师。

B)学区内分层抽样的50名教师。层或类是男性和女性,以及高中、初中和小学老师。

C)哈特高中的6名教师和詹姆斯高中的5名教师。

做如下提问:你是否有足够的数据计算得分、平均数和方差以外的东西? 例子A是可以的,但例子C则不太可能。

例子B怎样呢? 假设你正在调查一组教师,其中60%为女性,40%为男性。如果你抽取了一个50人样本并且你的分层方法是成功的,那么一个组中将有30名(60%×50)女性,另一组有20名(40%×50)男性。根据t检验所基于的统计理论,30和20的组刚好足够大。

考虑下面的表:

学　校	教师数	
	男	女
小　学	2	15
初　中	3	10
高　中	15	5
合　计	20	30

如果你希望比较小学的男女教师,那么你的分析中只有 2 男 15 女(尽管总共有 20 男和 30 女)。这就引出了在选择分析方法时需要提出的 7 个问题中的第 2 个。

2.你要考察关系或关联吗?

■ 示例

关系和相关

A.投票人的政治观点(在"政治观点调查"中表达出来的)和正规教育年限之间的关系如何?"政治观点调查"给出了 10(自由)到 100(保守)的连续分数。

B.在"政治观点调查"中的高分数是否与独裁—自由评估中的高分数相关联?中低分数又如何呢?

■

在第一个例子中,政治观点分数和正规教育年限这两个变量之间的关联程度,可以通过使用皮尔逊积矩相关系数计算得出,因为数据是连续的。对于第二个例子,具有高中低的等级,因此斯皮尔曼等级次序相关系数是合适的。

3.是否进行组间比较?

■ 示例

比较组

A.哈特高中的 15 名教师与詹姆斯高中的 20 名教师相比较。

B.来自学区高中、初中和小学的 100 名男女教师将在 12 月和 6 月接受调查并对两个时点进行比较。

■

如果你要做组间比较,卡方、t 检验、ANOVA、几率比或相对风险以及 McNemar 检验等统计方法有助于你确定观察到的差异是真实存在的还是偶然的。

4.你的调查将进行一次还是数次?这个问题的目的是确定调查设计是截面的还是纵向的。

■ 示例

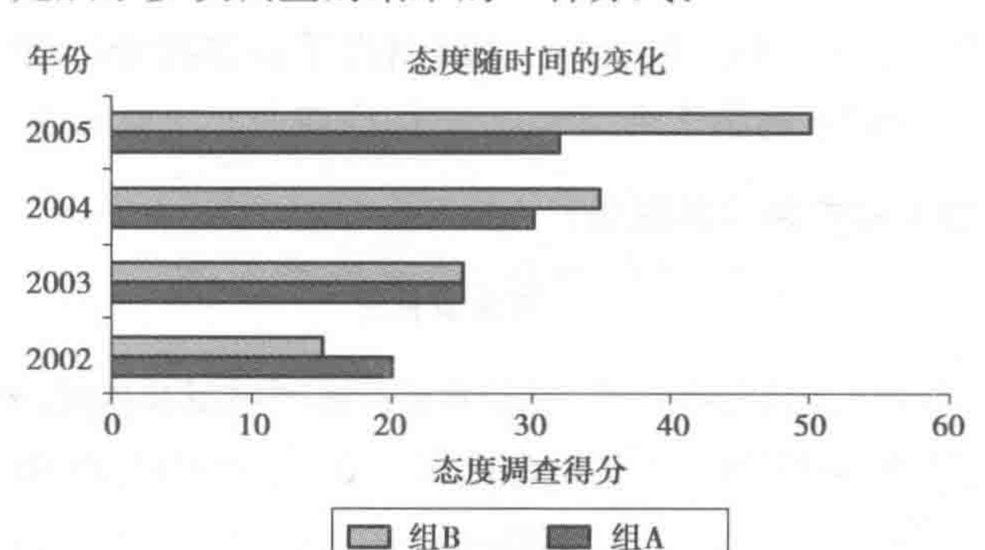
找出变化

A.所有 500 名教师将在 12 月被调查。

B.一个 100 名教师的样本将在 12 月被调查,然后在 6 月份被再次调查。

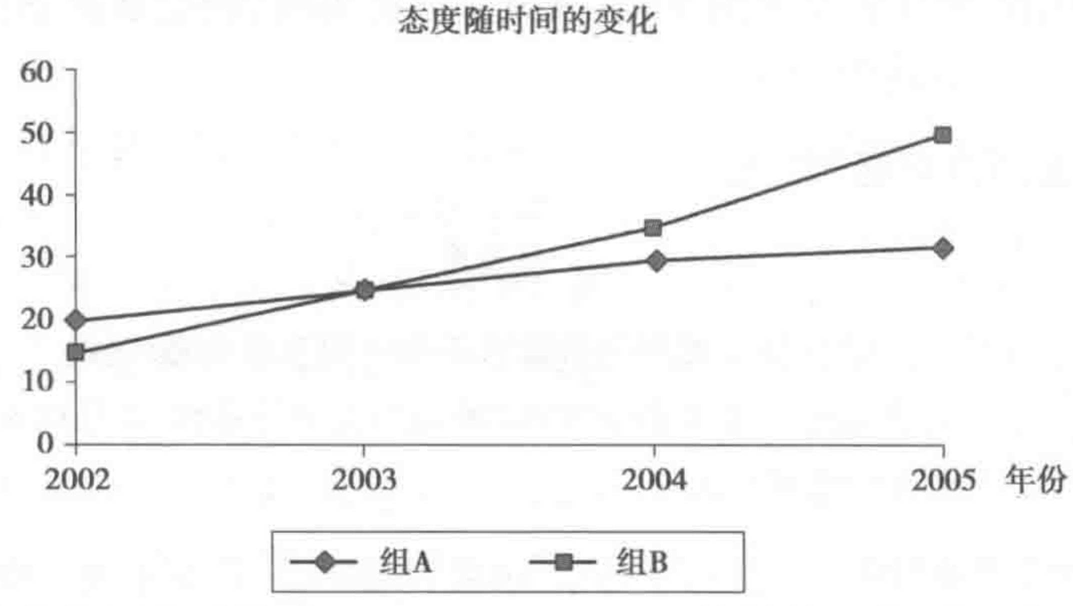
■

以下是展示多次调查的结果的一种方式:



图形显示在 2002 年,组 B 的分数低于组 A。两组的分数都随时间有所增长。组间的差异是否显著?不同时间的差异是否显著?该图不能告诉我们。对于统计显著性,你需要使用统计方法。

这里是另外一种展示历时性变化的方式,使用了相同的数据:



这张线图能让你直观地看出分数变化的坡度。但是与条形图一样,不能仅通过看图就能说出变化是否显著或者显著程度如何。

5.数据是数字和百分比还是分数和平均值?看下面:

A)一套调查数据包含同意或不同意学区教育政策表述的教师的人数和百分比。

B)一套数据包含对各种教育方法的态度(自由对保守)进行测量的得分。

调查数据可以是数量和百分比(分类数据)或者适合计算平均值

的分数(连续数据)。对两者使用的分析技术是不同的。参见表 6.1 和表 6.2。

表 6.1 比较法学院录取的男性和女性:人数

	辅导项目 A(n=102)n(%)	辅导项目 B(n=206)n(%)
男	50(49)	100(49)
女	52(51)	106(51)

表 6.2 比较法学院录取的男性和女性的分数:均值

	辅导项目 A(n=102)$\bar{X}$[1] SD[2]	辅导项目 B(n=206)$\bar{X}$ SD
男	20.6[3] 2.3	42.8 2.9
女	15.7 1.4	40.2 2.1

1.均值

2.标准差

3.分数的范围从 10=低到 50=高

比较来自不同辅导项目(分类数据)而被法学院录取的人数,与比较态度评估的平均分数(连续数据),需要不同的分析技术。例如,第一个比较可能会用到卡方,而第二个可能用到 t 检验。

6.你感兴趣的因变量和自变量有多少?

一个变量是你的调查感兴趣的一种特征。健康状况、对学校的态度、自我效能、组(实验、对照)、性别、工作满意度以及生活质量都是变量。自变量用来预测或解释研究发现。它们通常包括人口学特征(诸如年龄、民族、性别、收入和被调查者属于实验组还是控制组)。因变量是你的调查探究的东西:知识、态度和行为。

假设你想知道男性和女性的生活质量是否有所不同。自变量是性别(男和女),因变量则为生活质量。如果你想比较男女的生活质量和健康状况,你就有了 1 个自变量(性别)和 2 个因变量(生活质量和健康状况)。如果你提出下面的问题就意味着增加了一个自变量:实验和对照组中的男女在生活质量和健康状况上相比如何?增加的自变量是"组"(实验和对照)。

为了选择合适的分析方法,你需要知道自变量和因变量的数量。考虑下面的对话:

调查者:我想比较一下两组大四学生的生活方式偏好。每组有超过300人。

分析者:你所说的生活方式偏好是指什么?你打算比较倾向于某种生活方式——比如说农村与城市——的大四学生的人数?还是比较他们在一项调查中的得分?

调查者:两组大四学生都完成了LIFEQUEST,一个100题的调查。它产生从1到100的分数。较高的分数代表更精神化的生活方式,而较低的分数则意味着更物质化的生活方式。

分析者:如果你有连续的分数,那就是连续数据。对于两组和连续数据,t检验应该不错。

调查者:如果我打算比较人数的话用什么方法正确?

分析者:应该是卡方分析。

调查者:假设我有两组以上和连续数据呢?

分析者:那可以考虑ANOVA。

7.数据是高质量的吗?

如果调查是可信和有效的,针对的是正确数量的正确的人,并且记录准确,那么数据就是高质量的。如果数据是不完整和不真实的,那么用什么分析方法都无所谓,因为结果都是没什么意义的。

数据管理

怎样才能确保数据是完整的、可靠的,且适于分析、解释和报告?回答这个问题需要学习数据的组织和管理。

组织或管理调查数据是数据分析的一个基本部分。它是分析过程中将数据输入数据库以便获得结果的这部分内容。但是在输入数据前,你需要确定如何准备数据,以便以有用的形式获得你所需要的信息。

编写编码簿或操作手册

编码是对统计软件“说话”的基本单位。假设 5 200 名大学生完成了一个称作“大学酒精相关问题调查(CARPS)”的饮酒调查。你关注的一个问题是有多少男性报告每天一次饮酒 5 杯。为了找到答案，你需要告诉计算机去寻找哪个变量(如,性别和饮酒的数量和频率)。你还得告诉计算机如何区分男性和女性,以及怎样得到每人饮酒量和饮酒频率的数字。

计算机程序通过编码读取诸如性别和数量/频率这样的变量。看看下面这个来自 CARPS 的节选。

■ 示例　**节选自 CARPS——一个检测大学生酗酒的调查**

1.你是男性还是女性?［性别］

男性 $\square_0$

女性 $\square_1$

2.在过去的 12 个月中你多久一次饮酒 4 杯及以上?［QFDRINK］

选择一个答案

每天或几乎每天 $\square_1$

每周 4 或 5 次 $\square_2$

每周 2 或 3 次 $\square_3$

每月 2 到 4 次 $\square_4$

每月 1 次或更少 $\square_5$

从不 $\square_0$ ■

编码是右侧回答框下面的数字。你使用统计程序告诉你有多少人对问题 1 的回答是 0 且对问题 2 的回答是 1。为了做到这一点,你需要告诉计算机变量的名称(问题 1=SEX 和问题 2=QFDRINK)和值(0=男性,1=女性;以及 1=每天或几乎每天,…,0=从不)。

括号里的“词”(“SEX”和“QFDRINK”)是代表每个问题的变量。所有的编码簿都包含对与调查相关的问题、编码和变量的描述。编码簿必须围绕分析的目的来建立。看看下面节选自 CARPS 的编码簿。

■ 示例 节选自 CARPS(一个检测大学生酗酒的调查)的编码簿

变量名	变量标签	值:标签和编码
PROJID	项目编号	7 位 ID
AGE	年龄	生日:月/天/年
GENDER	性别	被访者性别:0=男性;1=女性;如果缺失空着
CNTRY	出生国家	出生的国家:01=阿根廷;02=玻利维亚;03=智利;04=哥伦比亚;05=古巴;06=厄瓜多尔;07=萨尔瓦多;08=瓜地马拉;09=洪都拉斯;10=墨西哥;11=尼加拉瓜;12=秘鲁;13=波多黎各;如无数据则空着
FGUILT	因为饮酒而感到负罪或内疚的频度	1=每天或几乎每天;2=每周至少 1 次,但少于每天;3=每月至少 1 次,但少于每周;4=少于每月 1 次;0=从不。如果无数据则空着
DRIVE	饮酒和驾车	1=在过去的 12 个月中 20 或更多天;2=10-19 天;3=6-9 天;4=3-5 天;5=1-2 天;0=从不;如果无数据则空着

■

所有的变量都被划分为分离的单元,称作值,与该变量的编码相一致。例如,因为饮酒而感到负罪和内疚的频率有 5 个值:1=每天或几乎每天;2=每周至少 1 次但少于每天;3=每月至少 1 次但少于每周;4=少于每月 1 次;0=从不;如果无数据则空着。编码是 4,3,2,1 和 0。如果没有来自被调查者的信息,则什么都不填。

虽然统计程序的术语各不相同,但几乎都要求以大写字母命名变量,并且避免使用特殊字符,如逗号或分号。有些程序对你所能使用的字符数(通常是 8 个)有限制。变量标签是变量的真实名称(如,"感到负罪和内疚的频率"是名为 FGUILT 的变量的真实名称)。为了理解你的数据,计算机需要知道变量名(如 FGUILT)、变量标签(感到负罪和内疚的频率)以及值标签(每天或几乎每天=1;每周至少 1 次

但少于每天=2,等等。)请注意:你的统计程序可能使用略微不同的术语,但意思都是一样的。

在大的调查项目中,编码簿是项目的官方记录。编码簿包含调查的测量工具;变量名、标签和值或编码;调查方法和特点;研究发现;以及调查测量工具(包括计分系统,如果相关的话)和调查团队成员。如果你或你的同事想使用调查数据(比如说做其他分析)或重复调查(比如针对不同的样本),就会发现编码簿是必备的。因为其详细,这种类型的编码簿有时也被称为操作手册。

示范性的完整的编码簿可以从网上找到。为此,使用你最喜欢的搜索引擎,输入"调查 AND 编码簿"或"+调查+编码簿。"

进行可靠的编码

为了确保小型调查——比如说只有一个人进行编码——的数据可靠,需要对所有数据或一个样本重新编码,以检查一致性。第二次编码应该在第一次编码一周后进行。这是为了留出充足的时间让编码员忘记第一次的编码,而不会去自动地重复。在数据二次编码后,需要对两个编码集进行比较。

在大型或超大型调查中,应该由第二个人独立地对数据的一个样本进行编码。为了确保编码员之间的可靠性,需要提供所有术语的定义并正式培训所有编码员。

尽管你为建立高质量的编码簿和数据管理系统付出了最大的努力,但编码员之间的不一致仍然会经常出现。为了找出他们的一致程度,即编码者间或评分者间信度,你可以计算一个称作 Kappa 的统计量。Kappa 测量的是一对编码员之间的一致程度能比偶然一致好多少。

测量一致程度

假设两个审核人被要求独立地对 100 个访谈进行审核,这 100 个访谈针对是否新建一个购物中心进行投票的社区成员。审核人检查访谈记录,找出有多少人在讨论中提及用全国连锁店取代单一业主店。如果没有至少提及一次这个问题,就要求审核人编码为 1,代表

“未提及”,否则编码为 2,代表“提及”。以下是审核人的编码。

		审核人 1		
		未提及	提及	
审核人 2	未提及	20[C]	15	35[B]
	提及	10	55[D]	65
		30[A]	70	

审核人 1 认为 100 个被访谈者中有 30[(A)] 人没有提及取代单一业主店,而审核人 2 认为有 35[(B)] 人没有提及。两个审核人都认为未提及的有 20[(C)] 人。

什么是描述审核人之间一致程度的最佳方式? 100 中的 20 或 20%[(C)] 可能低估了,因为两个评审人还都同意 55%[(D)] 的人提及了取代。把两者合起来(55%+20%)又高估了,因为只有两类(是和否),有些一致可能是偶然出现的。

这显示在下面的公式中,其中 O 是观察到的一致,C 是偶然一致。

测量两个编码者之间的一致:Kappa 统计量(κ)

$\kappa = O - C$(非偶然的一致)

$1 - C$(可能非偶然的一致)

以下是针对前例如何应用该公式。

1.计算有多少被访谈者是审核者因为偶然而一致认为未提及取代单一业主小店。做法是否定回答的数量相乘再除以 100(因为有 100 个被访谈者):$30 \times 35/100 = 10.5$。

2.计算有多少被访谈者是审核者因为偶然而一致认为提及取代,再乘以每个审核人认为提及的被访谈者人数。做法是将肯定回答的数量相乘再除以 100:$70 \times 65/100 = 40.5$。

3.把步骤 1 和步骤 2 中获得的两个数字相加再除以 100,就得到一个偶然一致比例:$(10.5+45.5)/100 = 0.56$。

观察到的一致是 $20\% + 55\% = 75\%$ 或 0.75。所以,非偶然一致是 $0.75 - 0.56 = 0.19$,这是分子。

可能非偶然一致是 100%−偶然一致的 56%或 $1 - 0.56 = 0.44$,这是分母。

$$\kappa = \frac{0.19}{0.44}$$

$$\kappa = 0.43$$

Kappa 多高算“高”？有些专家为 Kappa 值划定了范围：0.00-0.20＝极低；0.20-0.40＝较低；0.40-0.60＝中等；0.60-0.80＝较高；0.80-0.10＝极高。

怎样才能获得审核人之间较高或很高的一致——信度？可以通过确保所有的审核人在完全相同的题目下收集和记录数据，并且在每个重要变量的含义上事先取得一致。例子中 0.43 的 Kappa 值可能是因为在审核人和调查者的定义上存在差异、在使用定义上训练不足，以及编码错误等原因造成的。

检查调查的缺失数据

缺失数据来自调查中未回答的问题，或者干脆根本没有完成调查。你必须在收到第一批完成的调查后，开始录入数据前，就着手检查。在自填、书面调查中，特别是邮寄问卷中，容易出现未答问题。被调查者可能因为不想、遗漏（没看见）或不理解而未回答问题。人们可能因为内容太多或文字太深而不能理解调查的指导语或问题。被调查者可能不确定如何作答（完整地填空或圈选正确的答案，等等）。他们还可能觉得格式难以使用。

令人困惑的问题形式也会导致缺失数据，因为被调查者不知道如何回答问题。大量的认知前测和试调查可以部分解决被调查者误解调查题项的问题。认知前测指的是对潜在的被调查者进行访谈，请他们解释每个问题并回答。试调查是在计划的调查环境中对调查进行检测。这两者能够事先告知你哪种格式不可用或某些问题讲不通。请注意所有调查类型都必须进行试调查，包括商业性的网上调查。问题形式可能对某些被调查者而言不合适，调查可能包含不太相关的问题，网站难以使用，等等。然而，铁定的是，无论你怎么做，总会有缺失数据：这是肯定的。

调查者面临的一个主要问题是如何处理调查问题。比如说你寄

出 100 份问卷回收了 95 份。你自豪地宣称你的回答率为 95%。在经过仔细检查后,你发现半数人没有回答问题 5,并且这不是 25 个问题中唯一一个没有被所有调查者回答的问题。因为这些缺失信息的存在,你就不能宣称在所有问题上都获得了 95%的回答率。

对于缺失回答应该做些什么?在某些情况下,你能够回到被调查者那里,请他们回答遗漏的问题。在小型调查中,被调查者在哪里是知道的——比如说,办公室或学校里——可以很容易地联系上被调查者。但是对于大多数调查,再次收集信息就算不是不可能的也是不现实的。有些调查是匿名的,你甚至不知道被调查者是谁。在机构环境中,你可能需要回到机构审查委员会去获得再次接触被调查者的许可。这个过程可能会花费时间并延误你调查的完成。这些情况下,可以采用统计方法处理缺失数据。有一些是基于一般的被调查者将会如何回答的假设可以运用。

网上调查可以将程序设计为被调查者必须回答一个问题后才能进行到下一个问题。然而,有些被调查者可能觉得这种方式很不舒服,因而拒绝完成调查。虽然强迫被调查者回答所有问题被吹捧成网上和其他调查的一个主要优点,但有些调查者认为强制被调查者回答每个问题是强迫命令性的和不道德的。这些调查者认为,在许多调查中,参与者是自愿的,被调查者可以随时退出。一些迫使人们回答问题的调查,即使可能本意并非如此,也会被某些人认定违反了有关人身自由和尊重个人的道德原则。此外,这种做法可能会给调查带来不可信的信息,因为有些人可能会为了进入下一题而输入一些无意义的回答。

输入数据

数据录入是把调查的回答转换为统计软件可以读取的格式的过程。通过统计软件数据进入数据库。通常有三种形式。第一种,编码后的调查数据被输入数据库管理程序或电子表中。数据被存储为文本或 ASCII 文件,因而可以被输出 SPSS、SAS 或 Stata 等统计程序中。第二种数据录入类型是把数据直接输入 SPSS、SAS 或 Stata 等统计程

序中。在第三种数据录入形式中,被调查者或访谈人直接把回答输入计算机。这种类型的数据录入涉及计算机访谈、网上调查和扫描调查。回答被自动输入数据库管理系统或统计程序。还有程序能自动将一种文件格式转换为另外一种文件格式(比如说,从 SAS 到 Stata)。

每种数据录入、数据库管理和统计程序都有自己的惯例和术语。有些程序在你输入数据时会告诉你,你正在建立每个被调查者的"记录"。每个记录包含该人唯一的 ID(识别码)和该人的"观察"(回答选择、分数、评论等)。其他程序认为分析的单位(如个人)是观察,对每个观察所收集的数据是变量或字段。下面的例子是一个简单的 6 人数据集。

■ 示例

来自 6 人的调查回答

被调查者 ID	性　别	婚姻状况	孩　子	酗　酒	内疚感
1	1	3	2	1	1
2	2	4	1	3	2
3	2	3	2	3	2
4	1	3	2	1	2
5	1	1	1	1	2
6	1	3	1	3	1

■

在这个例子中,表格的组织是行代表每个人,列是其数据。也就是说,第 2 个人的数据是 2、4、1、3 和 2。被调查者 ID 是"被调查者识别码"的缩写。许多统计程序需要你告诉计算机变量的数据位于哪里。例如,在本例中,性别变量称作"GENDER",性别数据能够在第 2 列找到。酗酒变量称作"BINGE",这个变量的数据位于第 5 列。

数据管理程序、统计程序和具有自动数据录入的计算机辅助调查,可以通过只允许输入合法编码来提升准确性。例如,如果输入的编码必须为 001~010,那么你可以撰写一条规则不允许输入 01 或 10。如果你试图输入 01 或 10,就会出现错误提示。只需少量编程,程序就可以检查每项输入,确保与之前输入的数据相一致,以及跳转模式得

到遵守。也就是说,程序能够保证被某些被调查者跳过的问题字段编码为跳过而非缺失数据。设计计算机辅助规则需要技巧和时间。除非经过反复测试,否则不应该认为规则是无误的。

清理数据

一旦输入被录入,需要对它们进行清理。任何人使用一个"干净"的数据集都能得出相同的分析结果。数据变"脏"的原因有很多,包括误编码、不正确的数据录入和缺失回答。

避免脏数据的一个好的方式是使用扫描或网上调查。然而,如果你选择手工输入数据,就要确保编码员和数据录入员是有经验的、经过良好训练的和受到监督的。

当你录入了10%的回答时可以做个频数分析。反复做直到你确认调查运转正常。频数分析是对每个调查问题的回答的列表。如果你的数据集相对较小,你可以扫视频数查错。对于有许多表格或记录、变量、跳转和开放性文本回答的大型数据库,你可能需要做一个系统的计算机化的检查。所有主要的统计程序都提供清理规范,可以用于数据录入期间或之后单独的清理步骤。

还有其他几个问题要求你对数据进行整理。包括处置因问卷未返回而完全缺失的数据,以及包含异常值的问卷。

为了尽量避免低的回答率,要确保调查对于期望的被调查者是有意义的。题目是否重要到让他们感到有必要回答问题?要向被调查者表明你将遵守有关保密性的规则。通过认知前测和科学的翻译检查调查的可读性水平、格式和适合性。如果阅读水平过高、格式过于密集或翻译不佳,人们就会轻视调查。如果你正在进行一项访谈,那么需要付出时间和金钱培训和监督访谈者。人们会因为不理解或不喜欢调查或访谈者而拒绝接受调查。提供财务或其他奖励经常能鼓励被调查者作答。

验证调查数据

网上调查软件几乎总是能让你验证回答。例如,如果一个问题是

问日期,你可以决定日期的格式。假设你的调查是问一个人的生日,你可以对程序进行设置,让被调查者只能以月、日、年的形式记录数据。使用其他形式(比如说日、月、年)的人,将不被允许记录他们的回答。他们将会看到一个“无效回答”的提示,或计算机发出警告音,或者两者皆有。验证功能还能保证回答的一致性,这对定量分析而言是基本的。

调查软件的基本版本(有时是免费的)通常没有验证功能,只有在需要付费的高级版本中才有这样的功能。

一些经常能在网上调查软件中看到的验证指示和选择包括以下这些:

- 不要验证评论文本
- 必须以规定的长度
- 必须是一个整数
- 必须是一个分数
- 必须是一个日期(MM/DD/YYYY)
- 必须是一个日期(DD/MM/YYYY)
- 必须是一个电子邮件地址

小 结

调查数据分析使用统计和定性方法去描述、比较和解释被调查者的态度、偏好、价值和行为。

为你的调查选择一种合适的分析方法时,你必须考虑一系列因素。这些因素包括调查的样本量和研究设计,以及数据的特征和质量。

为了选择方法,回答下面 7 个问题:

1.你调查了多少人?

2.你要考察关系或相关吗?

3.是否进行组间比较?

4.你的调查将进行一次还是数次?

5.数据是数字和百分比还是分数和平均值?

6.你感兴趣的因变量和自变量有多少?

7.数据是高质量的吗? 也就是说,它们是可信的和有效的吗?

内容分析用于分析定性数据,旨在推论出被调查者做出的开放性回答和评论的意义。

要对所有返回的调查资料进行缺失数据的检查,并制订处理缺失数据的计划。

缺失数据来自未回答的问题或整个调查。

数据录入是把调查的回答转换为统计软件可以读取的格式进而输入调查数据库的过程。

网上调查数据能够被自动地输入数据库,并能实时地提供报告。调查数据还可以通过扫描或手工输入。

不管数据如何进入数据库,都要对回答的准确性进行检查。

干净的数据集是任何人使用了都能得到与你的分析一致的结果。网上调查软件通常具有“验证”选项,以标准化数据收集,并保证被调查者提供的信息就是调查询问的信息。

思考一下

1.找到下面列出的两篇文章。指出和讨论每篇文章用来解释其调查结果的数据分析方法。

你可以在任何搜索引擎中输入杂志名,通过 PubMed 找到这两篇文章。无须付费。

文章 1. O'Connor, T. M., Yang, S.-j., & Nicklas, T. A. (2006). Beverage intake among preschool children and its effect on weight status. *Pediatrics*, *118* (4), e1010-e1018.

文章 2. Richardson, J., Coid, J., Petruckevitch, A., Chung, W. S., Moorey, S., &Feder, G. (2002). Identifying domestic violence: Cross sectional study in primary care. *British Medical Jouranl*, *324* (7332), 274-278.

2.对于问题 1 中列出的两个调查(或其他调查)回答下列问题:

调查了多少人?

调查者是否确定了关系或做出了预测?

调查者是否进行了组间比较?

他们讨论的调查开展了一次还是数次?

数据是定类的? 连续的? 还是两者都有?

调查涉及的因变量和自变量有多少?

数据是高质量的吗? 也就是说,调查者是否表明了数据是可信的和有效的?



7

展示调查结果



调查结果可以用表格、图形、图解或图片的方式展示。

调查报告经常会有汇总表。其目的是描述被调查者，并展示关系与变化。如果你打算采用表格则需要遵守一些规则。例如，把自变量，诸如调查的组别和时间（如：基期、6个月后和3年后），放在表格的列上。选定一种表格形式并沿用之、谨慎使用竖线以及按逻辑顺序排列数据，如从最经常到最少见。

饼图展示每个回答类所占的比例。饼图的每个扇叶必须与其所代表的回答的数量或百分比成比例。

条形图经常被用于展示调查数据，因为它能够在一张图中提供多种信息——比较和变化。线图也用来展示变化和组间差异。注意不要过分夸张，使得1%的变化也在图上显得十分显著，除非确实如此。

许多调查结果以书面形式报告。考虑在报告中包括以下内容：摘要、总结、内容目录、图表、术语表、关于目的或目标的陈述、方法、结果、结论、建议、参考文献、索引、附录和致谢。

如果你打算随时修改，或者希望立刻获得听众的反馈，

幻灯片演示是最好的选择。另外一个选择是海报,可以向读者提供关于调查目的、方法、结果和发现的密集的和视觉化的说明。

复制问卷

调查结果可以以调查的形式或表格、图形、图解或图片的方式展示。见下面的示例:

■ 示例 **以问卷的形式报告结果**

在**过去的 4 周**内,你感到沮丧的时间有多长?

选择一个框		
□	所有时间	5%
□	大多数时间	3%
□	很多时间	8%
□	有些时间	42%
□	很少时间	32%
□	没有时间	10%

■

使用调查形式呈现结果的好处是能够让读者或听众看到问题及其回答选项。这种方法展示了调查的原始数据。它以未经任何分析和说明的方式提供数据。

网上调查系统实时地提供信息。随着回答的完成,系统自动地给你信息,诸如到目前为止完成调查的人数以及回答的合计。下面是一个来自 QuestionPro 的例子。

■ 示例

实时的网上调查统计

调查统计	
观看	409
开始	661
完成	143
完成率	21.63%
退出(开始以后)	518
完成调查花费的平均时间:2 分钟	

你的年龄属于下面哪个年龄组?

频数分析

	回　答	数目	%	20%	40%	60%	80%	100%
1	10 岁或更小	59	10.89					
2	11~15	63	11.62					
3	16~20	90	16.61					
4	21~27	71	13.10					
5	28~35	81	14.94					
6	36~45	59	10.89					
7	46~60	60	11.07					
8	60+	59	10.89					
	合　计	542	100					

关键分析		
均　值	4.410	主要事实 31.55%选择下列选项: 16~20 28~35 选择最少的选项(10.89%) 10 岁或更小
置信区间@ 95%	[4.226-4.594] n = 542	
标准差	2.185	
标准误	0.094	

■

使用表格

调查报告几乎总是包含表格。其目的是描述被调查者以及显示关系及变化。看一下下面这些带边框的表格。

■ 示例　　描述两所学校孩子的带边框表格

特　征	学校 1	学校 2
	n(%)	n(%)
年龄	____	____
平均分	____	____
阅读分	____	____
数学分	____	____
科学分	____	____

■

■ 示例　　比较两所学校孩子的带边框表格

特　征	学校 1	学校 2		
	n(%)	n(%)	t	p
年龄	____	____	____	____
平均分	____	____	____	____
阅读分	____	____	____	____
数学分	____	____	____	____
科学分	____	____	____	____

■

第一张表格罗列了两所学校具有某种特征(如年龄和平均分)的孩子的比例。第二张表格纳入了检验差异是否有意义的统计量。

表 7.1 是另外一个例子,显示了统计上显著的差异(以 1 个和 2 个星号表示)。该表还列出了来源:全国儿童健康调查,健康统计中心。当信息来源不明显时要将其列出。当信息来自报告者以外的其他人时它就是不明显的。还要注意,表格的脚注中对分数进行了解释。

表 7.1 四所学校中具有健康和不健康生活方式[a]的儿童

	健 康	不健康
学 校	n(%)	n(%)
Alameda(n=263)	140(19)	123(17)
Berkeley(n=152)	100(14)	52*(7)
Delacorte(n=227)	89(12)	138**(18)
Santa Inez(n=91)	45(6)	46(7)
合 计	374	359

来源:由健康统计中心全国儿童健康调查实施的自填调查

[a] 75~100 分为健康生活方式。26~74 分为不算健康又不算不健康的生活方式。1~25 分为不健康的生活方式。

* $p=0.003$ 介于健康和不健康之间

** $p=0.002$ 介于健康和不健康之间

再看表 7.2 和 7.3。

表 7.2 自我效能的变化(%)

	小于 65 岁		65 岁及以上	
	2012	2013	2012	2013
水平	(n=128)	(n=104)	(n=49)	(n=212)
高	0	70	0	69
中	57	28	57	31
低	43	2	43	0

来源:自我效能量表(纽约:国家出版社)

注意:量表是从 1 到 9,其中,1~3 分=高;7~9 分=低。

表 7.3 自我效能的变化(%)

	2012		2013	
水 平	小于 65 岁 (n=128)	65 岁及以上 (n=104)	小于 65 岁 (n=49)	65 岁及以上 (n=212)
高	0	70	0	69
中	57	28	57	31
低	43	2	43	0

来源:自我效能量表(纽约:国家出版社)

注意:量表是从 1 到 9,其中,1~3 分=高;7~9 分=低。

哪张表的信息量更大,表 7.2 还是表 7.3? 如果调查强调的是组间(小于 65 岁与 65 岁及以上)的变化,那么表 7.2 更好。如果你想强调经历了 1 年时间以后发生的变化,那么表 7.3 更好。

一些制表的规则

1.表格以行列的形式显示数字、百分比、分数和统计检验结果。根据可读性确定表格的行列数。

2.每张表都应该有一个标题,说明其目的和内容。

3.当表格数据的来源不是那么显而易见时,应该列出。

4.当你使用一个容易引起误解的术语时,界定它。以星号或上标标示定义。

5.列为自变量,如调查的分组和时间(如基期、6 个月后和 3 年后)。

6.选定一种表格形式并沿用之。前例中的表格列中文字首字母大写(除了介词和冠词)。在行中,只有第一个字母大写。谨慎使用竖线,特别是当你已经使用了横线时。

7.按某种逻辑顺序排列数据。一种经常使用的顺序是从最经常的到最不经常的,不过相反的顺序也是可以的。要点是要有内在逻辑,以方便读者理解。

8.包含样本量以及数字、百分比和其他统计量的区分。

所有的字处理软件都能帮助你格式化表格使其易读。如果你经常使用表格,学一学如何使用你的软件的表格功能。

画饼图

饼图直观地展现了每个回答类占总体的比例。假设你开展了一个 80 名图书馆用户的需求的调查,并且想区分不同年龄用户的需求。再假设你发现 40 名被调查者相对年轻——比如说,介于 18~25 岁,而只有 10 人年龄在 25~35 岁,其他 30 人的年龄超过 45 岁。你可以通过图 7.1 有效地描述你的发现。

再看图 7.2。这张饼图有与之前相同的用户比例,但通过切开饼,强调了扇叶的不同大小。

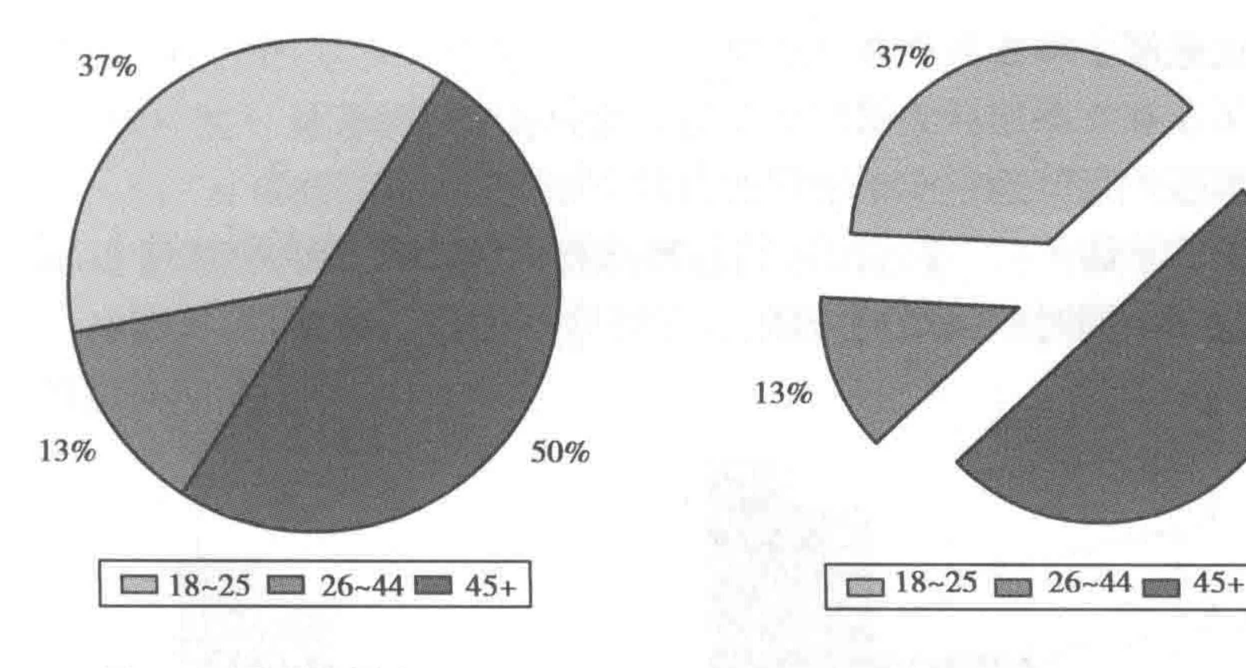


图 7.1 图书馆用户:$N=80$　　图 7.2 图书馆用户:$N=80$

如果你打算经常使用饼图,可以采用作图软件。包含作图功能的电子表、字处理软件和其他程序也能帮助你根据数据表画出饼图。准确的饼图的关键是其比例尺。每个扇叶必须与其所代表的回答的数量或百分比成比例。也就是说,50%的回答是半个饼,25%是四分之一,等等。记住扇叶的数量最好别超过 6 个,否则饼图会显得过于混乱。

使用条形图

条形图经常被用于展示调查数据,因为它能够在一张图中提供多种信息。参看图 7.3,一张关于男孩和女孩从 1985 年到 2005 年的行为改变的图。

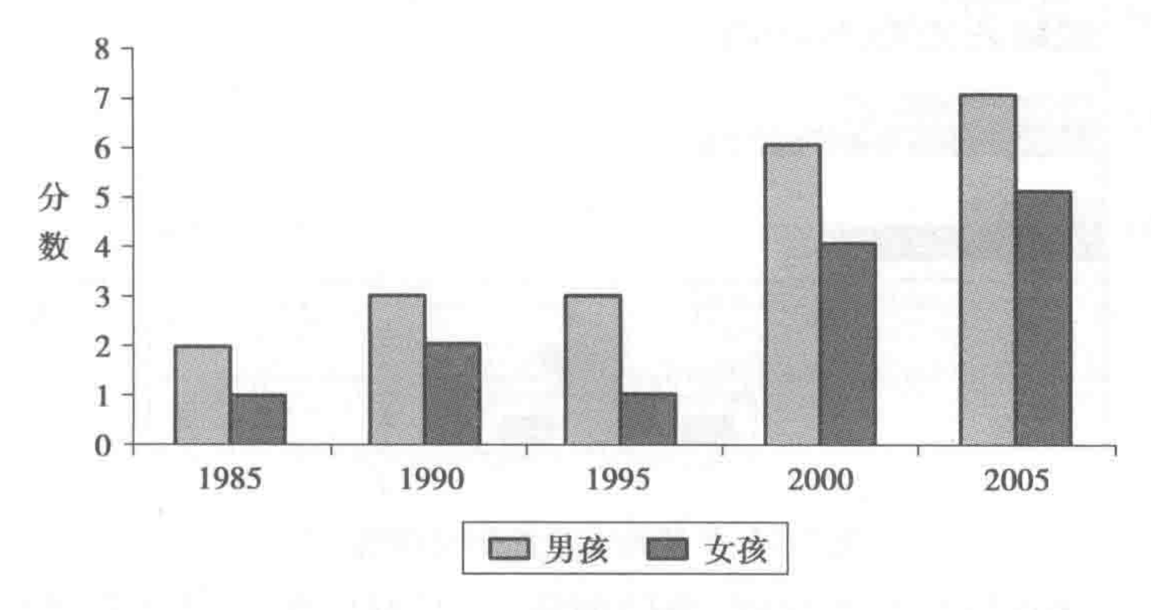


图 7.3 男孩和女孩的行为调查

这张图立即告诉你两件事情:

男孩和女孩的行为随时间发生了变化。

男孩的得分总是高于女孩的得分。

现在看图 7.4。表达相同信息的这张图集中在男孩和女孩上,虽然你仍然能同时获得两种信息。

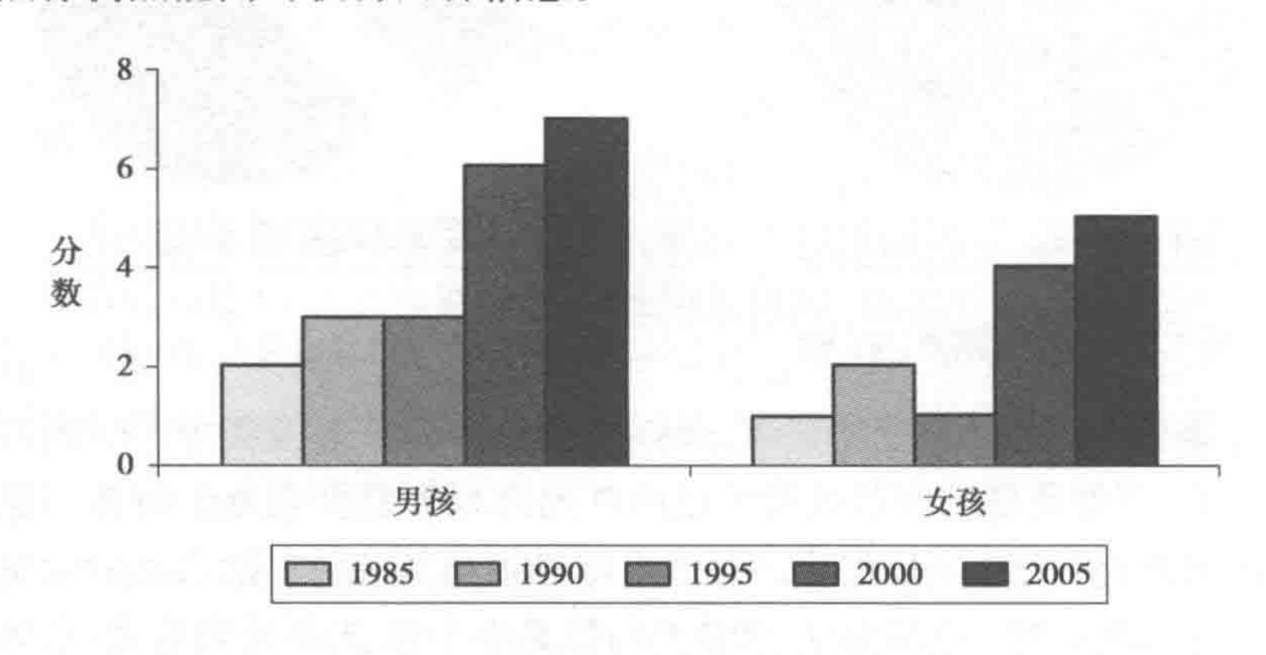


图 7.4　男孩和女孩的行为调查

图 7.5 是一个相同信息的图集中在分数上,你可以立即看到两类数据。

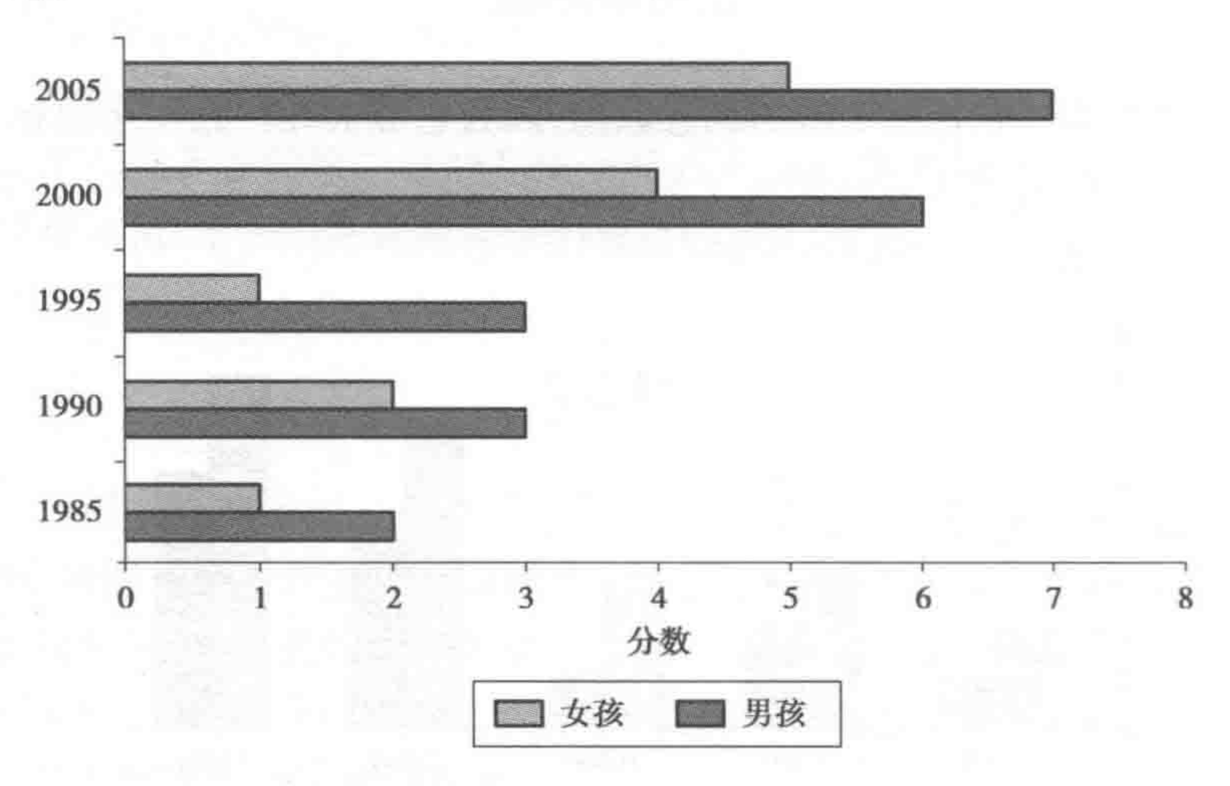


图 7.5　男孩和女孩的行为调查

条形图应该总是有标题、条的图例,以及有助于如实反映结果的

说明。记住眼见不一定为实。这是因为图形上的差异不一定是真实的差异(统计显著的和实际的)。

图 7.6 以线图展示相同的数据。线图能够比条形图更好地反映随时间变化的趋势。大多数作图程序能让你自动从一类图形转化为另一类图形,因此你可以实际看到哪种图形更适合描述你的调查数据。目标是准确展示。

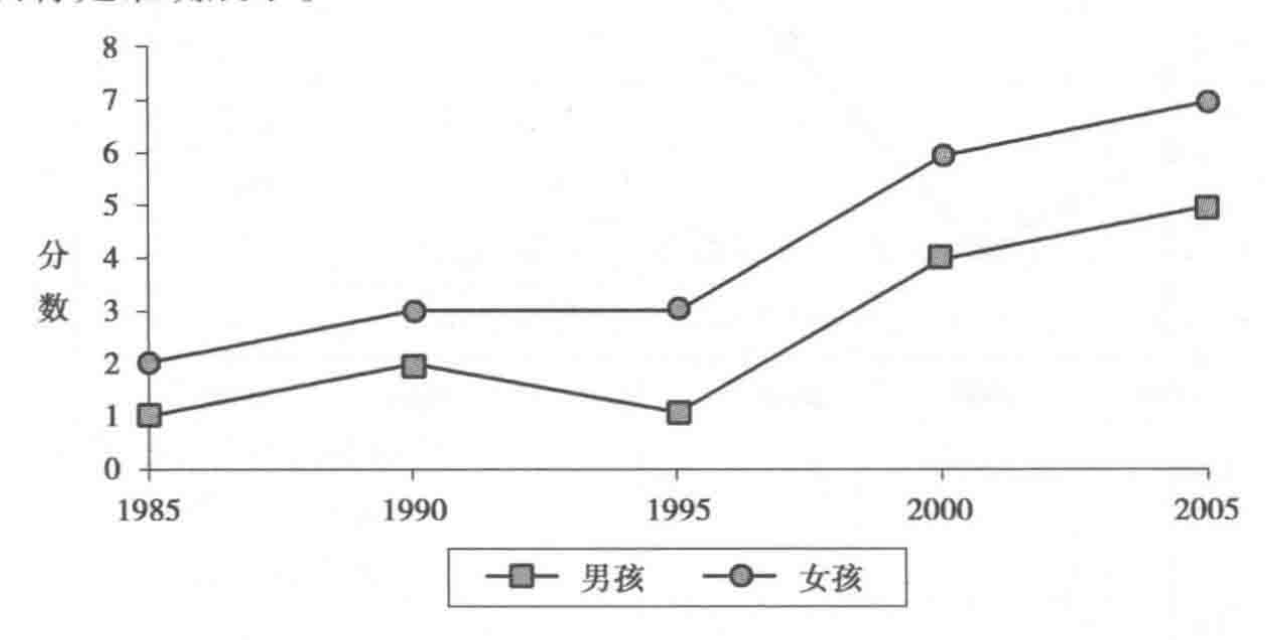


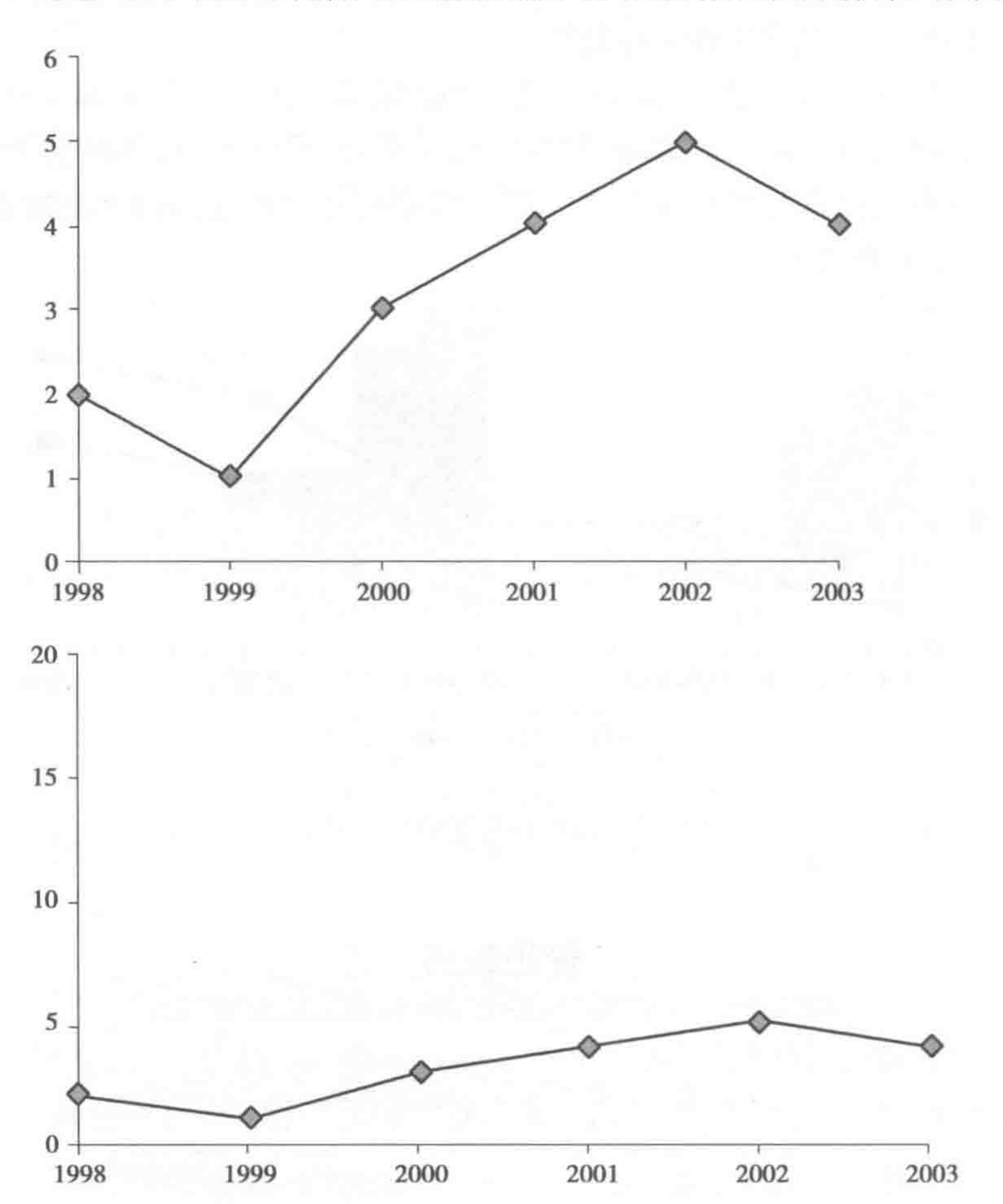
图 7.6 男孩和女孩的行为调查

使用线图

线图能让你展示变化和组间比较。注意不要过分夸张,使得 1% 的变化在图上也显得十分显著,除非确实如此。看下面这些分数:

被调查者人数	谁得到这一分数
1	2
2	1
3	3
4	4
5	5
6	4
合　计	19

这里有两个图形,都是以线图的形式反映分数及其频数(人数)。



哪个图更准确?第一张图是计算机的选择。第二张图不是。在第二张图中,得分的差距看上去没有那么显著。

哪个更可取,条形图还是线图?如果有异议,那就做实验。为了展示比较,表和条形图可能更有效。对于随时间的变化,线图较好。

网上调查系统有个好处是可以让你快速探索最佳的方式展示结果。下面给出了一个关于公司支持满意度调查的结果的不完整的例子。接下来是对一个顾客满意度调查的网上调查系统的报告的说明,并用条形图和线图描述了结果。

调查者通过把调查的回答导入一个电子表程序并使用其“向导”

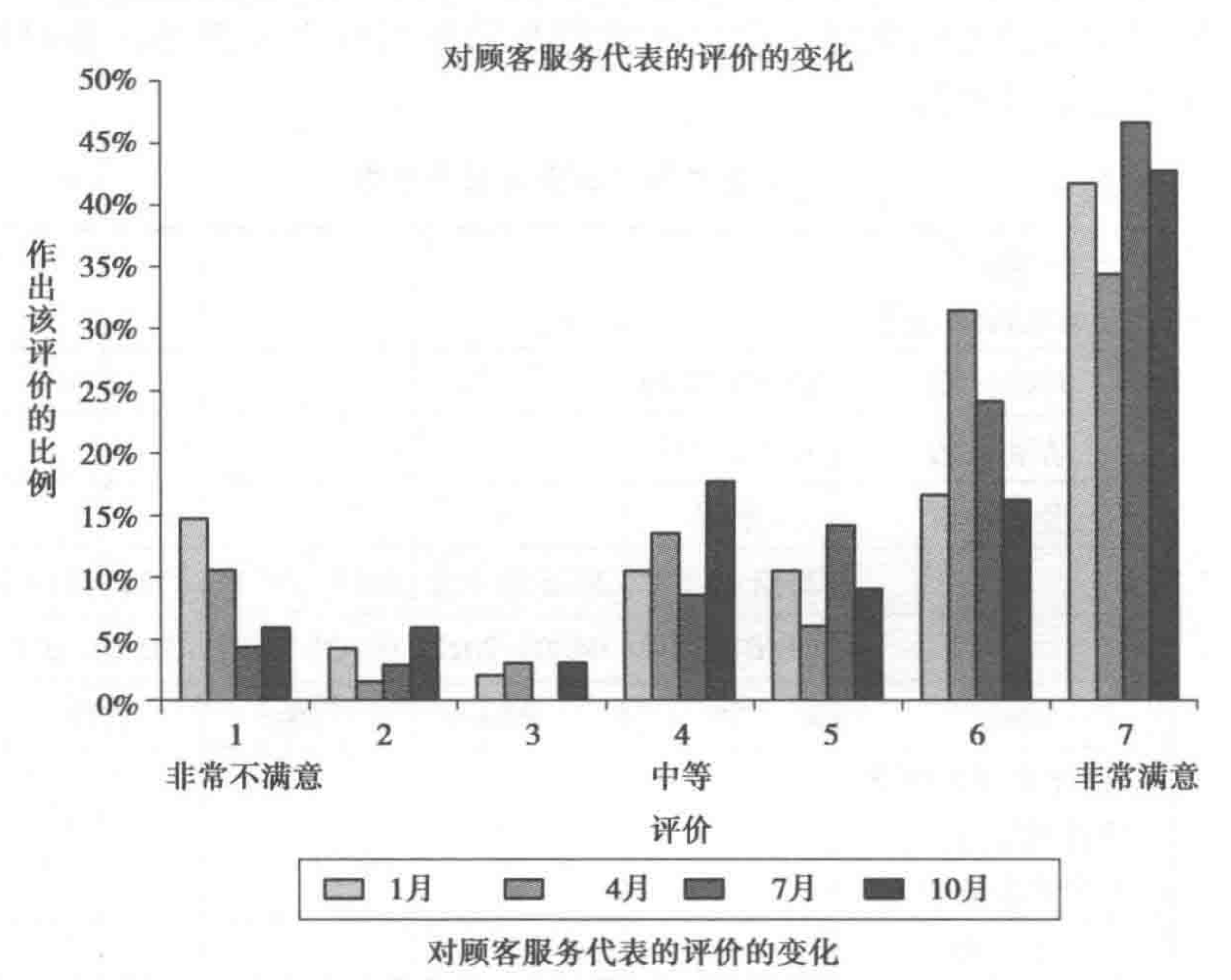
功能作图。条形图在观察每个评级随时间的变化方面是有用的。线图在展示变动方面更好。应该注意的是彩色图形在屏幕显示和打印出来时更容易理解。

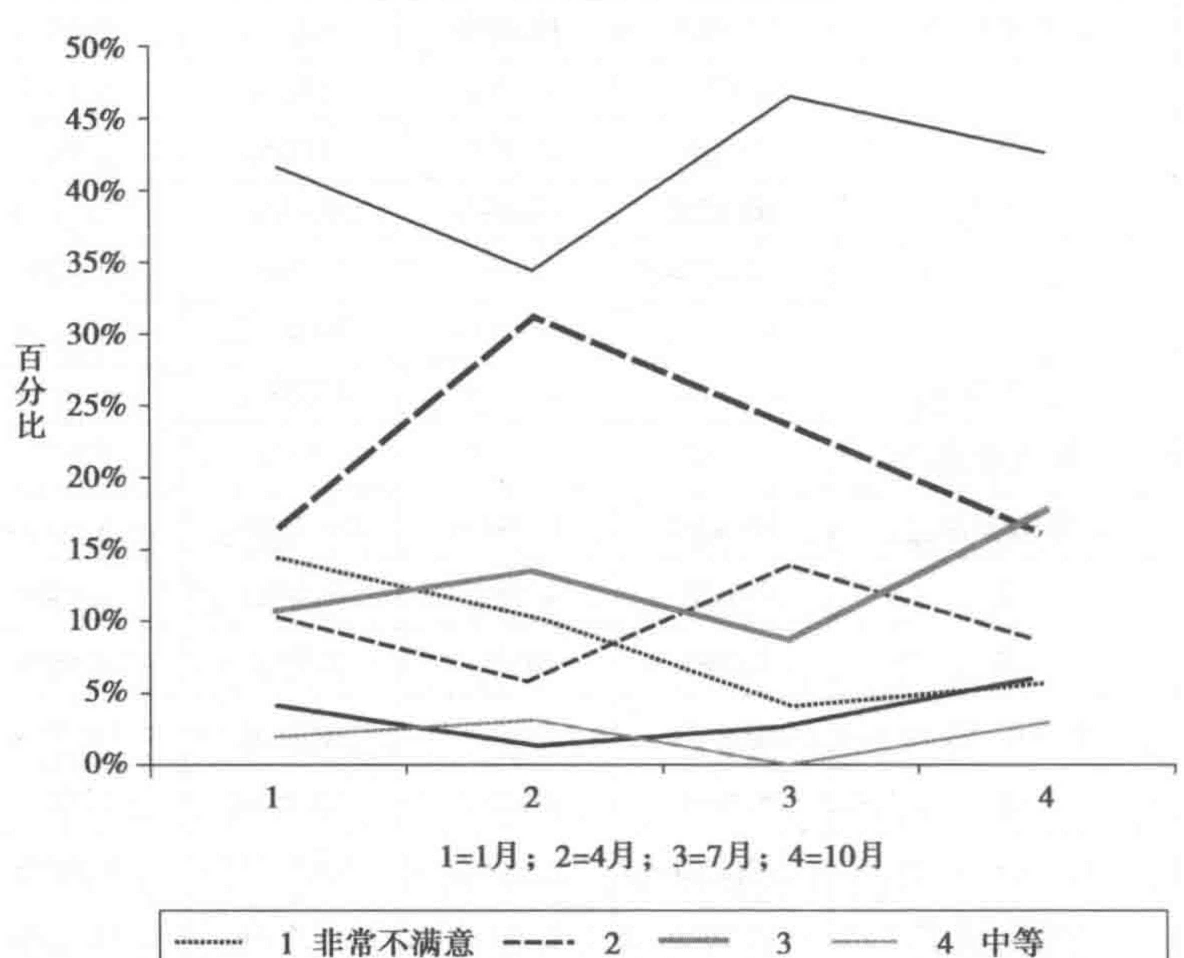
■ 示例　　**下载的顾客满意度调查结果**

趋势报告：公司支持满意度				
开始日期	01/01/2012			
结束日期	12/31/2012			
报告类型	季报			
	2012 年 1 月	2012 年 4 月	2012 年 7 月	2012 年 10 月
	01/01/2012	04/01/2012	07/01/2012	10/01/2012
	56	68	76	73
1.请评价你的顾客支持代表在下面5 个方面的表现				
专业性				
1 非常满意	14.58%	10.45%	4.23%	5.88%
2	4.17%	1.49%	2.82%	5.88%
3	2.08%	2.99%	0.00%	2.94%
4 中等	10.42%	13.43%	8.45%	17.65%
5	10.42%	5.97%	14.08%	8.82%
6	16.67%	31.34%	23.94%	16.18%
7 非常满意	41.67%	34.33%	46.48%	42.65%
帮助意愿				
1 非常不满意	16.33%	11.94%	4.35%	5.97%
2	6.12%	4.48%	2.90%	4.48%
3	2.04%	4.48%	2.90%	4.48%
4 中等	10.20%	8.96%	10.14%	10.45%
5	8.16%	10.45%	14.49%	7.46%
6	12.24%	22.39%	24.64%	20.90%
7 非常满意	44.90%	37.31%	43.48%	41.79%

来源：www.QuestionPro.com. ■

■ 示例　　把一个问题的结果转化为条形图和线图





■

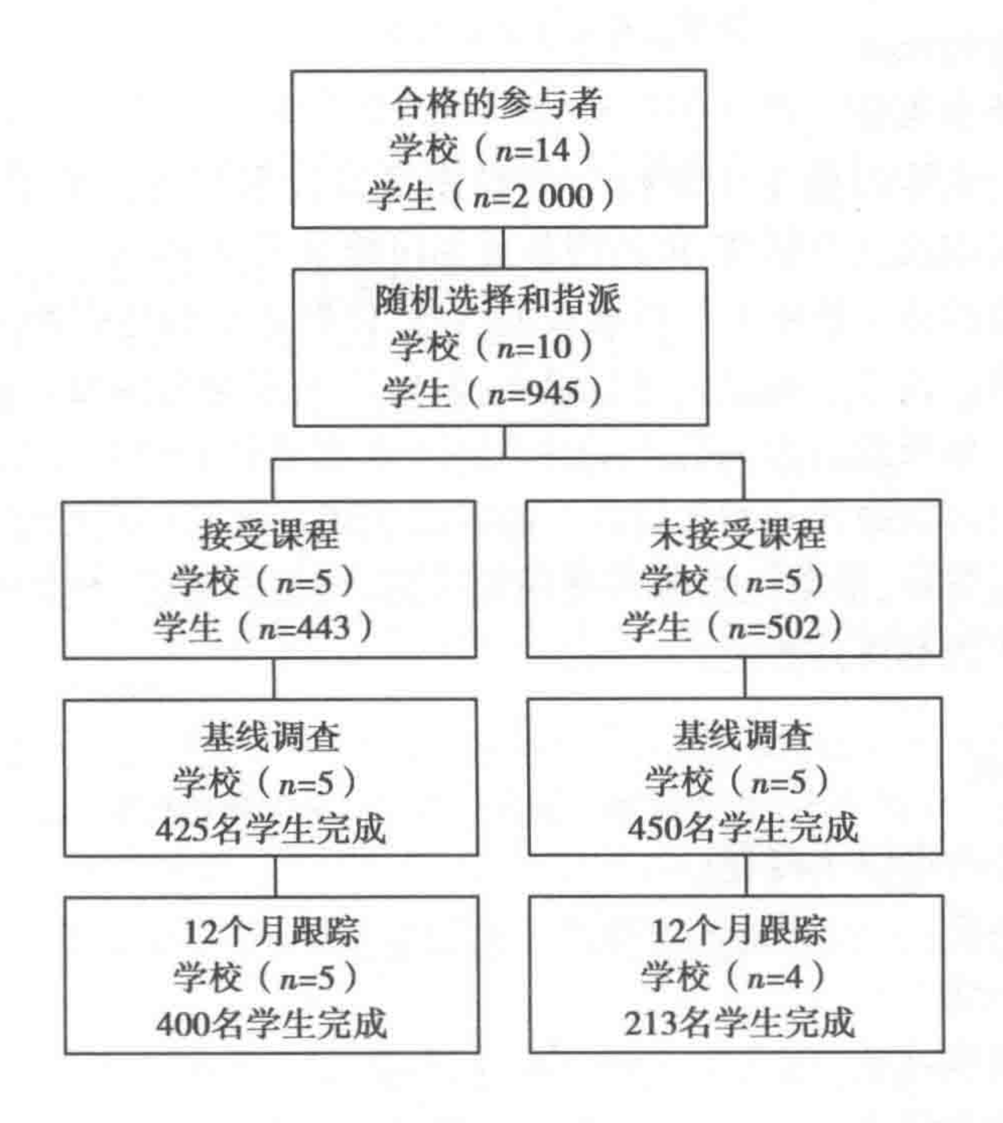
使用图示或图片

使用图片或图示说明你的意思。假设你想描述你的调查研究设计。比较文字与图示的区别。

■ 示例　　调查报告中的文字与图示

我们的设计以学校为抽样单位。14 所学校的总共 2 000 名学生有资格参加一个伤害预防项目。我们随机选择 10 所学校,这 10 所学校总共有 945 名学生。5 所学校及其 443 名学生被随机指派接受课程。在接受课程的 5 所学校中,425 名学生完成了基线调查。在没有接受课程的 502 名学生中,450 人完成了基线调查。我们获得了接受课程的 5 所学校的 400 名学生 12 个月的跟踪数据。有一个没有接受课程的学校中途退出了,我们收集了剩下的 4 所学校的 213 名学生 12 个月的跟踪调查数据。

现在看下面的图示。



■

类似于此的图示有助于解释复杂的研究设计。适用于组织或流程图的原则也适用于图示。

所有主要的字处理程序和电子表格都能让你很容易地把数据表格转化为图形。如果你经常使用流程图,你应该考虑购买一套作图软件。

撰写调查结果

调查活动的一个相当典型的成果是一份关于其目的、方法、结果、结论及建议的书面报告。当你坐下来组织报告的时候似乎总会遇到一个问题:应该写得多详细和多专业?如果太专业的话,可能会大幅度降低可读性;然而,忽视专业细节也容易让你的报告受到批评。

以下三个标准可以指导你的写作:

- 尽量完整
- 仔细组织
- 清晰表述

看上去像以前某个学校老师对你撰写读书报告的要求?可能。但如果采纳这三个标准,你的报告更加可能具有可读性。

完整性意味着纳入尽可能多的内容,以便让任何感兴趣的人都能理解报告。你必须对报告进行组织,以便让人方便地找到他或她想找的东西。如果我对你为什么选用"XYZ 自我效能调查"感兴趣,我应该能够很容易地找到相关信息。清晰的表述永远是一个期望的目标。在调查报告中,清晰性意味着使用标准的、规范的语言,并且确保所有模糊的术语都得到界定。

组织报告

需要考虑以下内容:

摘要
小结
内容目录
图表目录
术语表

线路图:在报告中能够找到什么以及到哪里查找
对目的或目标的阐述
后果
方法
结果
结论
建议
参考文献
索引
附录
致谢

摘要。摘要通常需 150~250 字。可以考虑采用结构化摘要。如下:

■ 示例　　　　**一个调查报告的结构化摘要**

目的:考察社区对把男孩和男人整合进 Women Forward(WF,一个传统上全部由女性组成的民间组织,其目标在于鼓励女孩和女人参与政治)的态度。

研究设计:对 150 人的截面调查

调查:2004 年 7 月开展的 15 分钟电话访谈。

被调查者:获得了 82%的回答率。被调查者中 70%为男性,62%的被调查者为 35 岁及以下。被调查者是从当地政治组织中随机选取的。

主要后果:对整合的支持

结果:84 个男性(97.6%)和 32 个女性(86%)支持整合。两性之间的差异不显著。35 岁及以下的被调查者比更年长的被调查者较不热衷于把男性整合进 WF 组织。

结论:虽然男性和女性同样支持把男性整合进 WF 组织,但年轻的被调查者无论性别意愿都较低。对这个发现要谨慎对待,因为 70%的被调查者是男性。 ■

这个摘要有 152 个字,包括标题。你的字处理程序可能有字数统计功能。

小结。小结是报告主要内容(目标、方法和发现)的精华。它的篇幅应该在 3 页纸左右,阅读时间在 5 分钟。你可以提供比在摘要中更多的细节,并提供谁开展的调查的信息。你还可以纳入 1~2 个重要的

图或表。

内容目录。内容目录应该列出报告的所有主要章节及相应页码。

图表目录。列出每个图表及其完整的标题和出现的页码。在有些报告中,会把图表放在最后单独的一节中。

术语表。所有专业术语(如随机抽样)、简写(如:CBS=商业和科学中心)以及模糊的概念(如:健康的、聪明的、敌意的)都应该包括在词汇表中。

线路图:在报告中能够找到什么以及到哪里查找。有些报告是复杂的,读者需要线路图的引领。例如,假设你的报告有两个主要的部分:第1部分和第2部分。告诉读者怎样找到每个部分(单独装订?点击一个词或一个图标?)。说明每一部分的内容(如:第1部分是对调查项目的一个描述,第2部分讨论了检验问题)。

对目的或目标的阐述。目的也就是调查的目标。在报告中讨论你为什么开展调查。你作调查是为了找出参与者对其收益的满意程度吗?还是为了找出哪种收益最令其满意?

后果。调查关注的主要因变量或者后果是什么?它们可能是知识、态度、行为、健康或生活质量的改变。

方法。做了什么?对谁?所付出的努力值得吗?为了回答这些问题,讨论下面这些内容:

- 调查的类型(书面的、自填问卷、电话访谈,等等)以及由此带来的局限性。例如,你对×县的青少年开展一项电话调查,你可以使用固定电话联系他们。你用固定电话是因为你能方便地从公共电话簿获得号码。然而,×县的青少年几乎都更愿意使用手机。如果你坚持使用固定电话,这会对调查的可信度产生何种影响?
- 调查所提出的问题。给出例子,如果正文不能容纳,可以把调查放在附录中。
- 调查安排。何时、持续多长时间以及多久开展一次调查?由谁来执行?执行者得到了怎样的培训?如何监控执行质量?由此带来的局限性是什么?例如,你培训了5个人,但其中1个不称职。这会对结果有何影响?
- 调查建构。给出原始的问题,描述试调查情况,讨论信度和效度。讨论建构期间产生的局限性。例如,验证性研究可能显

示该调查不适用于英语或西班牙语阅读能力不足9级的人。

• 抽样和回答率。入选标准是什么？你是怎样计算回答率的？回答率是否足够高？是否使用了奖励？讨论其局限性。如果没有获得足够数量的被调查者，对结果会有什么影响？例如，你希望有半数男人和半数女人，但是如果被调查者主要是女性呢？

• 调查研究设计及其局限性。例如，假设你的调查使用了非随机构成的对照组，预计会产生哪些选择偏倚？对研究发现的影响程度如何？

• 分析。针对每个主要关注的调查后果，描述分析方法。

• 伦理。是否正确告知了被调查者的权利和责任？是如何提及保密性的？

结果。对于每个感兴趣的后果，你发现了什么？通常以统计数字给出答案。

结论。研究发现意味着什么？

建议。基于数据你会给出什么建议？警告：不是每次都需要建议。有些人开展调查只是为了获取数据，他们会自己提出建议。

参考文献。引用文中使用的文献。按照字母顺序或者在文中出现的顺序排列文献。

索引。只需在长报告中使用索引。

附录。任何相关材料如果过于冗长的话都放到报告的附录里。也就是说把较长的调查放在附录中比放在正文中好。作为一个经验法则，15个问题及以下的问卷可以放在报告中。需要注意的是某些单个的问题可以有多个选项，这种可能需要放在附录中。例如，“在过去的12个月中，你是否有下列健康问题：头疼、胃疼、扭伤脚踝或手腕、感冒？”每个选项都等于一个单独的问题：“在过去的12个月中，你头疼过吗？”“胃疼过吗？”“是否曾经扭伤过脚踝或手腕？”“感冒过吗？”

致谢。感谢那些为调查的设计及其说明提供指导的人。需要感谢的还包括调查的参与者、统计分析和数据收集人员。

清晰写作指南

以下是如何清晰地写作的一些指南。

1.尽可能使用主动语态。

不好:报告是相对简单的,并且明显是针对非专业者的,因为只提供了很少的统计表格(20 个字,有"to be"形式的动词)。

较好:这个相对简单的报告明显是针对非专业者的,因为统计表格很少(17 个字,2 个动词)。

避免使用"there is"和"there are"

不好:报告中只有(用了 there are)很少的统计表格。

较好:报告中的统计表格很少。

2.不要用介词短语断开句子。可能时,应避免:"为了给一个理由"或类似表述。把介词短语转化为分词:"在获得可靠态度测量的努力中"可以变为"为了获得准确的态度测量。"把介词短语转化为形容词短语:"这是个具有重要性的问题"变为"这是个重要问题。"

将下面这些短语	替换为
at an early date	soon
at the present time	now
in order to	to
prior to	before
subsequent to	after

几乎所有的字处理软件都有语法功能。该功能能够很好地找出被动语态,还能提供替代选项,并提醒你可能忘记的语法。然而,它们并不是完美的,而且它们不关心写作的风格。

3.尝试某种可读性公式。可读性测试能够预测书面材料的年级水平。许多人阅读低于其年级水平的材料才会感到舒适,会忽视高于其水平的材料。所有主要的字处理程序都可以为你报告的全部或部分提供可读性统计量。

这里有个简单的公式可以用来估计可读性。应用这个公式得出的结果称作 FOG 指数:

①从你的调查报告中抽取一个 100 单词的样本。

②计算每个句子的平均单词数。如果样本中最后一个句子的单词数超过 100,使用到该句末尾的总单词数计算平均值。

③计算 100 单词样本中两个音节以上的单词的数量。不要

把专有名词或以 ed 或 es 结尾的三音节单词计算在内。

④每个句子的平均单词数量加上包含至少 2 个音节以上单词的数量,再将和乘以 0.4。

假设一个 100 个单词的段落每个句子平均有 20 个单词,10 个 2 音节以上单词,则和为 30。用 30×0.4 得到 FOG 指数为 12 年级。

看一下本章前面的结构化摘要。可读性统计量显示该摘要的平均难度要求阅读水平在 8 到 10 级之间。大多数可读性公式只关注单词。它们不包括数字。如果你的对象是外行,尽可能使数字简单。使用 76%或 76.2%,而不是 76.17%。

口头陈述

口头陈述的规则很多与书面报告相同。清楚而缓慢地说,控制节奏以符合听众的需求。在口头报告中你必须对报告的内容和重点进行选择,通常这是一件困难的工作。确保给听众有用的结果。政策制定者希望了解底线而理论家希望了解有效性。外行听众主要关注研究发现和参与者。如果有时间的话,你可以告诉他们调查的方法和局限。所有听众都希望听到调查问题的例子。

幻灯片演示

使用下列指导原则准备幻灯片。

- 一张幻灯片只包含一个主要概念。

不佳:

帮助学童应对暴力
- 抑郁症状
- 焦虑症状

治疗环境
- 家庭
- 学校
- 诊所
- 其他医疗环境

更好:

帮助学童应对暴力

- 抑郁症状
- 焦虑症状

治疗环境

- 家庭
- 学校
- 诊所
- 其他医疗环境

• 每张幻灯片留给听众 1~2 分钟。例外是当你拥有一系列相似的幻灯片时,如列表、饼图或图形。

• 如果你从教科书或其他调查中改编了信息,检查一下看是否需要简化。检查是否有版权方面的规定。

• 每张幻灯片的文字不要超过 8 行,每行不要超过 6 到 7 个单词。每个概念最多不要超过 10 个单词。不要指望听众也是读者。

不佳:

人头税暴乱

□大约 200 000 人游行到伦敦的特拉法加广场,这是反对一种称为人头税的新税制的全国性运动的一部分。

□示威和平开始,但却演变为 100 年来最大的暴乱之一。

□调查团队以示威者的身份参与到人群中,以此身份开展半结构化的访谈,并用手持录音机录下示威的声迹。

更好:

人头税暴乱

□200 000 人在特拉法加广场游行

□100 年来最大的暴乱之一

□目标是反对新税制

□示威和平开始

调查方法

□对示威者的半结构化访谈

□手持录音机捕捉“声迹”

• 为了强调要点，可以标题加下划线、使用要点或选择符号、给每个要点编号或对不同要点使用对比色。

• 句子和短语的使用上保持一致。

不佳：

> **真实验具有什么特征？**
> • 比较一或多个组
> • 有一个干涉
> • 随机选择
> • 随机指派人到组

评论：第一个要点符号后面跟着一个代名词（"-ing"）短语，而第二个后面是一个句子。第三个项目符号后是一个短语，第四个是句子。这种混杂容易使听众迷惑。

更好：

> **真实验具有什么特征？**
> • 一或多个组被比较
> • 受到一个干涉
> • 参与者是被随机选择的
> • 参与者是被随机指派到组的

这张幻灯中的每条都表述为一个句子。

• 如果存在"空白时间"没有合适的幻灯片可用，可以用过渡幻灯片。过渡幻灯片通常是一张透明的空幻灯片或者包含演示题目的幻灯片。不要使用动画作为过渡幻灯片，因为这会导致演示过程中的分心。（如果你在演示过程中使用了动画，要确保观众能够看清说明文字；还要确保动画使用没有版权限制或费用。）

• 使用分发的材料提供汇总信息和技术性细节及参考文献。确保你的姓名、演示的名称和日期出现在材料的每一页上。不要在你完成演讲前分发材料，除非你在演讲中会用到它们。

• 总的来说，大小写字母要比全是大写字母易读。

• 把数字四舍五入到最接近的整数。尽量避免小数，如果必须有的话，入到最近的十分位（32.6%，而不是 32.62%）。

• 把表格限制在 5 行和 6 列以内。

- 如果使用了坐标图,确保 x 轴和 y 轴被清楚地标识。
- 确保屏幕上所有的信息都在演示中得到讨论。
- 小心不要用动画、图形和声音压垮听众。警告!不要认为可以随便从网上下载图形或其他资料。其中相当一部分是有版权保护的。
- 每张幻灯片的颜色不要超过 4 种。如果你对所用的颜色不肯定,考虑选用程序预设的幻灯片演示颜色。品蓝背景的黄色或白色字母易于观看和阅读。
- 演示前检查幻灯片。确保字体是易读的和一致的。不要在一张幻灯片中混用不同的字体,除非是为了强调某些单词。如果可能的话,让其他人帮助检查一下幻灯片的内容和形式。

口头与书面报告:表达的差异

口头报告和书面报告至少在一个重要的方面有所不同。口头报告依赖于你讲述细节的能力。书面报告则靠文字和图片说话。比较下列两表:

■ 示例　　用作书面和口头报告中的表格

书面报告

表:基期和跟踪期均分(X)和标准差(SD)

	实验学校 (n=467 名学生)		对照学校 (n=400 名学生)				
调查后果	基期	跟踪期	基期	跟踪期	净差值	t	p
	X(SD)	X(SD)	X(SD)	X(SD)			
道德价值	75.6(11.8)	85.5(8.8)	78.8(10.9)	81.2(9.6)	7.5	8.9	0.000 1
宗教信仰	3.5(0.7)	3.8(0.7)	3.7(0.7)	3.8(0.7)	0.19	4.7	0.000 1
社会责任	3.7(0.7)	3.9(0.7)	3.7(0.7)	3.8(0.7)	0.10	2.2	0.003
伦理行为	1.5(2.5)	1.3(2.3)	1.0(2.0)	1.3(2.4)	−0.48	2.8	0.06

■

对表格的书面说明

该表给出了四个调查后果的基期和跟踪期均值,以及观察到的分数净差值。我们使用独立样本 t 检验比较实验组和对照组从基期到跟

踪期的均分变化。在道德价值、宗教信仰和社会责任上发现了有利于实验组的显著差异。社会责任上的差异虽然显著但没有达到 $p<0.05$——选定的 t 检验显著水平。

对表格的口头说明

甚至表格的外观也有所不同,它被简化了。

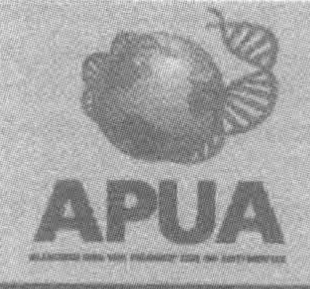
该表比较了实验和对照学校(指到相应的列)。我们使用独立样本 t 检验比较从基期到跟踪期的均分差异(指到相应的行)。我们发现在道德价值和宗教信仰上支持实验组学生的显著差异。社会责任上的差异达到显著但未达到 $p<0.05$,选定的 t 检验显著水平。

	实 验		对 照	
后 果	基 期	跟踪期	基 期	跟踪期
道德价值	75.6	85.5	78.8	81.2
宗教信仰*	3.5	3.8	3.7	3.8
社会责任*	3.7	3.9	3.7	3.8
伦理行为	1.5	1.3	1.0	1.3

* 基期和跟踪期之间存在统计差异。

海 报

海报是展示调查结果的一种重要方式。海报是被设计用来在没有口头说明的情况下总结你的工作的方式。使用相对大的输出页面,你可以使用简短的文字段落并以要点方式给出结果。大多数调查海报也包含图形元素帮助说明研究的重点。这里是海报的一个例子。它的真实尺寸是 36 英寸(91.44 厘米)乘以 49 英寸(121.92 厘米)。



Poster # 1 Patient Beh
Regarding
Implications f

Stephanie Boy
Alliance for the Pr

Background

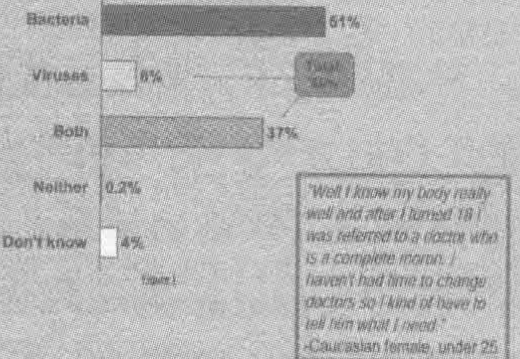
Antibiotic resistance is a growing problem which threatens the ability of health care providers to provide effective treatments for infectious diseases. Many patient behaviors contribute to the development of resistance, including obtainment of antibiotics from sources other than health care providers, the practice of stopping antibiotic prescriptions early, and pressuring of health care providers to prescribe antibiotics for conditions for which they are inappropriate[1,2]. The purpose of this research was to gain a better understanding of the beliefs and risk factors associated with such behaviors, and to identify ways in which clinicians can use this information to more effectively discourage inappropriate antibiotic use patterns in their patients.

Methods

In September through November of 2006, a sample of 919 English speaking U.S. adults who had taken antibiotics within the last 12 months was interviewed by telephone. Respondents answered questions about their practices, experiences, and beliefs concerning antibiotic use. Responses were weighted to reflect the U.S. population by gender, ethnicity, and age based upon the U.S. Census Bureau's 2005 Population Estimates. All differences noted between segments are significant at the 95% confidence level.

Results

Survey results indicate that, while many patients believe they "know what they need" to treat illnesses, knowledge on basic issues related to antibiotic use is often quite low. 10% of respondents identified Tylenol, Robitussin, or both of these as antibiotics, and an additional 4% named other non-antibiotic drugs as antibiotics at some point in the survey. Additionally, nearly half of respondents reported a belief that antibiotics are useful for treating viruses. (See Figure 1).



"Well I know my body really well and after I turned 18 I was referred to a doctor who is a complete moron. I haven't had time to change doctors so I kind of have to tell him what I need."
-Caucasian female, under 25

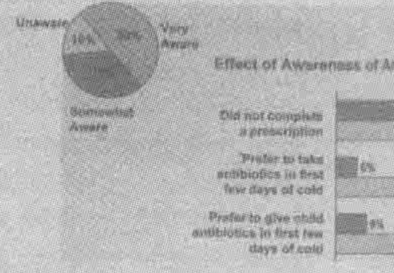
Results (continued)

More specifically, 8% of respondents stated that antibiotics
can treat viral illnesses more than doubled the chances tha
preferences, and summarizes the reasons respondents ga
antibiotic was a belief that it would speed recovery. This p
for their children. 13% of parents preferred to give their ch
can treat viral infections was stronger, quadrupling prefere

"[I knew I needed antibiotics] because I was having migraines."
-African American male, under 25

"It's common knowledge that antibiotics help you to get over a c
-Caucasian Female, 35-44 years old

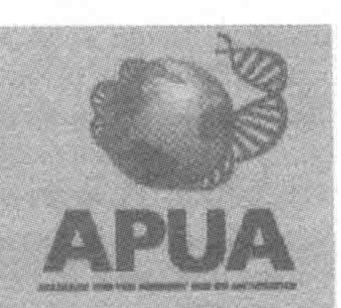
Knowledge about the types of illness that antibiotics can ef
early cold. Additionally, behavior was found to be mitigate
also reduced the likelihood that a respondent would report



Reported expectations for antibiotics were also strongly af
previously noted for bronchitis[3]. Figure 4 illustrates this ef

"I just took them because the symptoms seemed ve
similar to some of the illnesses that I had where an
antibiotic was prescribed. I just basically had some
of virus that just wouldn't go away. I had Amoxicillin
treat that in the past. It worked in the past."
-White male 25-34 years old

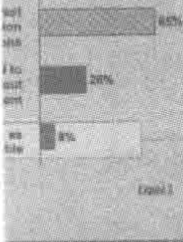
aviors and Beliefs
Antibiotic Use:
or Clinical Practice

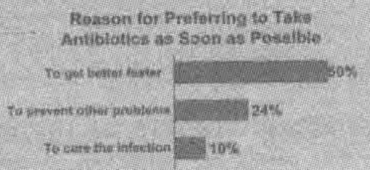


d, MA; Susan Foster, PhD
lent Use of Antibiotics, Boston, MA

e their first choice of treatment for an early stage cold, and a belief that antibiotics
respondent would state this preference. Figure 2 illustrates cold treatment
for preferring antibiotics. The most common rationale for preferring to receive an
ern was even stronger among parents asked about their cold treatment preferences
ren antibiotics right away, and in this group, the effect of believing that antibiotics
for antibiotic treatment.

eferred Response to
Cold in First Few Days

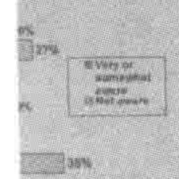




or a virus faster."

tively treat lowered the chances that a respondent would prefer an antibiotic for an
y having knowledge about the existence of antibiotic resistance. This knowledge
pping a prescription early without consulting a health care provider. (See Figure 3).

iotic Resistance

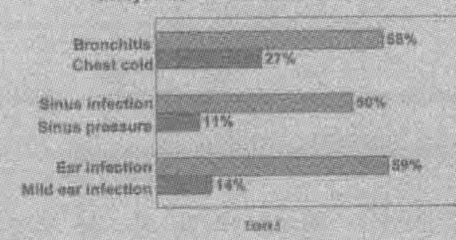


"I don't believe they are particularly good for you. I know that sometimes you need to take them. I think that many people are given too many antibiotics and they build up this immunity so then you have these superbugs."
-Caucasian female, 45-54 years old

"I read up on antibiotics and I heard that they kill the good cells as well as the bad ones."
- African American Male, 35-44 years old

ted by the terminology used to communicate diagnosis, an effect that has been
for three conditions that are often cited as being over-treated with antibiotics.

Expectation that Illness Should "Almost Always" Be Treated with Antibiotics



Conclusions and Implications for Practice

These results suggest potential communication strategies which nurse practitioners can use to reduce patient expectations for antibiotics in situations in which they are unnecessary. By understanding potential gaps in patient knowledge and communicating diagnosis in ways that patients find less threatening, clinicians can more effectively negotiate encounters in which patients exert pressure to prescribe an unneeded antibiotic. Specifically, the results indicate that the following practices may be useful:

1) Remember that patients may not have sophisticated understanding of what antibiotics can do, and take this into account when talking with patients who have requested unnecessary antibiotics. For example, a statement like "you have a virus, so you don't need an antibiotic today" might not be accepted as a logical argument to the 46% of survey respondents who said that antibiotics can treat viruses.
2) It is also important to remember that patients usually have specific reasons for their preferences. Since the most common reason given for preferring antibiotics for a cold was "to get better faster," it may be worth emphasizing the fact that antibiotics cannot help this happen.
3) Similarly, when giving a patient a prescription, it may be helpful to make sure the patient understands what type of medication he or she is receiving.
4) Survey results show that having knowledge about antibiotic resistance does indeed lead to lower rates of patient antibiotic requests and other undesirable behaviors. It is worth the time to explain to patients how resistance develops, and how this could directly affect them.
5) Communicating diagnosis in terminology that patients find less threatening may substantially reduce expectation of receiving an antibiotic.

Literature Cited

1) Welschen, I, et al. (2004). Antibiotics for acute respiratory tract symptoms: patients' expectations, GPs' management and patient satisfaction. Family Practice, 21(3), 234-7
2) Scott, JG, et al. (2001). Antibiotic use in acute respiratory infections and the ways patients pressure physicians for a prescription. Journal of Family Practice, 50 (10), 853-8.
3) Phillips, TG and Hickner, J. (2005). Calling acute bronchitis a chest cold may improve patient satisfaction with appropriate antibiotic use. Journal of the American Board of Family Practice, 18(6):459-63.

Acknowledgements

APUA gratefully acknowledges the members of its advisory board for their contributions to the design and analysis of this survey:

Timothy Edgar, PhD, Emerson College
Cindy Friedman, MD, CDC
Ralph Gonzales, MD, University of California, San Francisco
Gordon Grundy, MD, Aetna (retired)
Philip Watson, MD, University of Cincinnati



APUA also thanks Links Media fort he fielding of this survey and compilation of survey results and analysis.

海报应该提供一个关于研究的综述,包括对某个调查或某些调查的描述。它不是一篇学位论文或杂志文章,所以不要试图囊括所有的细节。随便一个观看者应该能在3~5分钟内掌握大致信息并在10分钟读完所有文字。

海报包含一篇杂志文章的标准部分。这包括:对研究目标、方法(包括研究设计、抽样和数据分析)、结果和结论的介绍。还可以加入摘要、致谢和参考文献。

海报应该有一个能在25英尺外看清的主标题。人们会经过一排海报,所以你需要在一定距离外抓住他们的眼球。一般规则是标题使用72号字,以及像Times New Roman或Arial这样常用的字体,并使用稍小一些的相同字体作为小节的标题。使用一种简单的颜色方案。不要因为使用过多的颜色、字体和字号而让人们分心。

在你设计之前就要考虑好如何打印海报。因为并不是每个打印选项都提供相同的纸张尺寸,并且较大的海报打印的花费也更高,所以首先选定打印的纸张尺寸,再开始设计海报。检查你的打印机零售商是否有某些特别的限制或指导。

Microsoft PowerPoint是一个相对易于使用的创建海报的工具。Adobe Illustrator有更多功能并能提供非常专业的成品,特别是对包含很多高分辨率图像的海报,但它也更为复杂和昂贵。稍微搜索一下,可以找到免费的(当然也有收费的)创建海报的网上模板。例如,某家网上公司有一个免费的PowerPoint(.ppt和.pptx原生格式的)研究海报模板集。你可以下载合适的PowerPoint海报模板,添加文字、图像和图形,然后将其发送回公司供打印。

小 结

使用调查形式提供结果的优点是能够让读者或听众看到问题及回答选项。网上调查系统基于实时提供信息。

每张调查统计表都应该有一个标题,说明其目的和内容。

当表格的数据来源不是一目了然时,应该给出来。

如果你在表格中使用了一个容易引起混淆的术语,要界定它。使

用星号或上标字母将其与定义关联起来。

表格列是诸如调查的组别和时间(如:基期、6个月后和3年以后)一类的自变量。

饼图展示了每个回答类占总体的比重。

图形图常用来展示调查数据,是因为它能同时提供几类信息的综合。

线图是可以让你展示变化和比较组的图形。

尽可能在调查报告中使用主动语态。

不要用介词短语分割句子。

对你报告的内容尝试使用可读性公式进行测试。

每张幻灯片只包含一个主要概念。

每张幻灯片留给听众1~2分钟。

为了在幻灯演示中强调要点,可以标题加下划线、使用要点或选择符号、给每个要点编号或对不同要点使用对比色。

总的来说,大小写字母比全是大写字母易读。

海报是一种以视觉化方式呈现调查情况的方式。海报的内容与研究论文类似,但是应该能让读者在10分钟之内掌握大概信息。

思考一下

1.为本章内容准备一个幻灯片演示。

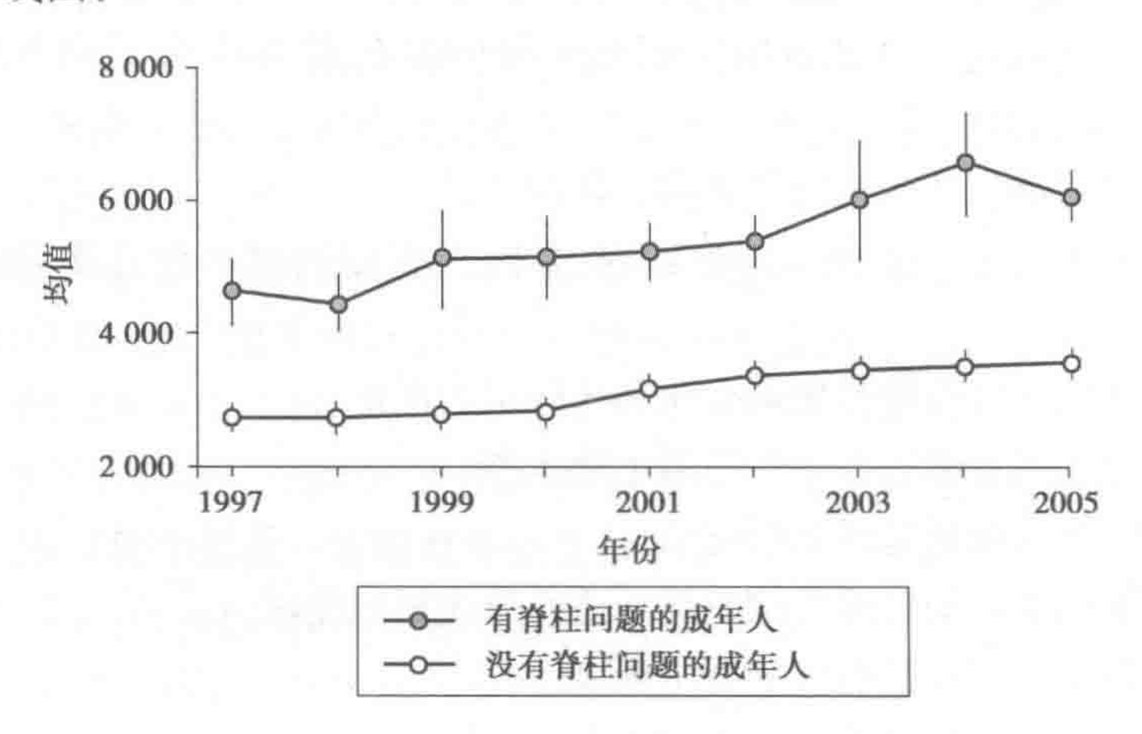
2.写一段话描述下面表格和线图中的调查结果。

表格:

有1个及以上现役士兵的家庭虐待孩子的情况

	虐待孩子的父母的数量(%)(n=1 858)			
发生虐待的天数	1个孩子被虐待	2个孩子被虐待	≥3个孩子被虐待	合计
1	1 028(55)	404(22)	241(13)	1 673(90)
2	41(2)	73(4)	49(3)	163(9)
≥3	0	7(<1)	15(1)	22(1)

线图：



3.描述你将如何在口头演示中提交上面表格和线图中的结果。书面报告和口头报告有何相同之处？有何差异？

为了建设好“万卷方法”，更好地服务学界，凡购买重庆大学出版社万卷方法系列图书的读者，填写以下信息调查表（复印即可），邮寄给我们（401331 重庆市大学城西路21号重庆大学出版社 林佳木），经过认证后，将获赠经管之家（原人大经济论坛）论坛币100个。

读者情况调查表	
姓名	
单位	
联系电话	
E-mail	
经管之家ID	
购买书籍	
购买渠道	
对书的评价或丛书的建议	



经管之家

——国内最大的经济、管理、金融、统计类在线教育网站

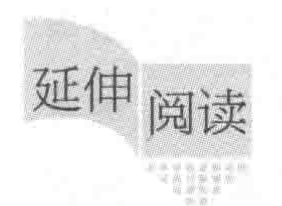
经管之家（http://www.jg.com.cn，前身为人大经济论坛），于2003年成立，致力于推动经济学科的进步,传播优秀教育资源,做最好的经管教育。目前已经发展成国内最大的经济、管理、金融、统计类在线教育和咨询网站，也是国内最活跃和最具影响力的经济类网站。

1.拥有国内经济类教育网站最多的关注人数，注册用户以百万计，访问群体遍布高校、行政机关和企事业单位。

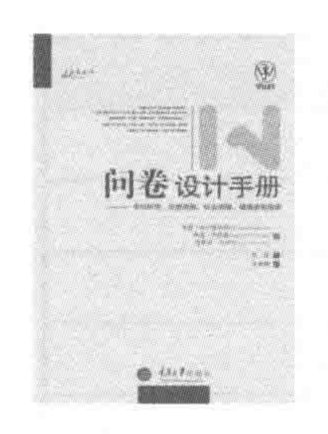
2.是国内最丰富的经管类教育资源共享与发布平台：各类经管书籍网络资源能最大满足您的要求；丰富的经济金融数据库使您不再为数据发愁；更有全面的经管统计类培训和教学相关软件。

3.拥有完善的教辅系统，主要为经管类教师和学生提供备课和学习资料，包括案例库、题库和期刊信息大全。

只要您是学习、研究或从事经管类行业，经管之家就能满足您的需求。







01 问卷设计手册：市场研究、民意调查、社会与健康调查指南

根据问卷的不同用途，将原理和应用技术结合讲解，对具体的问卷设计工作有良好的针对性。



02 管理学问卷调查研究方法

全面讲解了问卷调查研究几乎所有方面的知识和技能，包括文献回顾、理论发展、假设检验，以及目前被最广泛运用的问卷数据统计分析方法。



03 社会调查设计与数据分析：从立题到发表

全书分为 15 章，逐步讲解确立题目、调查设计、数据收集、数据分析的各种统计方法和适用范围，最后提供论文写作的模版提纲。以解析作者本人发表的国际期刊论文的方式，分享经验。

{
  "filename": "MTM5ODQwNzQuemlw",
  "filename_decoded": "13984074.zip",
  "filesize": 25952245,
  "md5": "75e190817ad9c8018d9fd933c5d872a5",
  "header_md5": "c83ff8abcd20d702ed2f5724c2a49886",
  "sha1": "7061b0c8f6e928edd6547b1a40fb7d8ce2068a28",
  "sha256": "368dacf39b5f64411992d53a93e9a451a9719ef92e139866492478c28c350bda",
  "crc32": 2498749147,
  "zip_password": "wcpfxk&*^TDwcpfxk",
  "uncompressed_size": 33364547,
  "pdg_dir_name": "",
  "pdg_main_pages_found": 218,
  "pdg_main_pages_max": 218,
  "total_pages": 232,
  "total_pixels": 932917104,
  "pdf_generation_missing_pages": false
}